

Evolution of Precipitation Structure During the November DYNAMO MJO Event: Cloud-Resolving Model Inter-comparison and Cross-Validation using Radar Observations

By

Xiaowen Li, Morgan State University/NASA Goddard Space Flight Center

Matthew A. Janiga, University of Miami*

Shuguang Wang, Columbia University

Wei-Kuo Tao, NASA Goddard Space Flight Center

Angela Rowe, University of Washington

Weixin Xu, Colorado State University

Chuntao Liu, Texas A&M, University at Corpus Christi

Toshihisa Matsui, ESSIC, University of Maryland, College Park/NASA Goddard Space Flight
Center

Chidong Zhang, NOAA Pacific Marine Environmental Laboratory

*Current affiliation: University Corporation for Atmospheric Research (UCAR) Cooperative
Programs for the Advancement of Earth System Science (CPAESS)

Key Points:

- Multi-radar observations allow stringent multi-model comparisons and validation for precipitation structures during an MJO event
- Cloud-resolving models constrained by site-specific observed large-scale forcing can reproduce variations of radar echo-top heights
- The multi-radar, multi-model framework also benefits quantitative sensitivity studies to highlight common model strength and weakness

A manuscript submitted to the Journal of Geophysical Research
September 2017, revised December 2017

Accepted February, 2018

Abstract

Evolution of precipitation structures are simulated and compared with radar observations for the November Madden-Julian Oscillation (MJO) event during the DYNAmics of the MJO (DYNAMO) field campaign. Three ground-based, ship-borne, and space-borne precipitation radars and three Cloud-Resolving Models (CRMs) driven by observed large-scale forcing are used to study precipitation structures at different locations over the central equatorial Indian Ocean. Convective strength is represented by 0-dBZ echo-top heights, and convective organization by contiguous 17-dBZ areas. The multi-radar and multi-model framework allows for more stringent model validations. The emphasis is on testing models' ability to simulate subtle differences observed at different radar sites when the MJO event passed through. The results show that CRMs forced by site-specific large-scale forcing can reproduce not only common features in cloud populations, but also subtle variations observed by different radars. The comparisons also revealed common deficiencies in CRM simulations where they underestimate radar echo-top heights for the strongest convection within large, organized precipitation features.

Cross-validations with multiple radars and models also enable quantitative comparisons in CRM sensitivity studies using different large-scale forcing, microphysical schemes and parameters, resolutions, and domain sizes. In terms of radar echo-top height temporal variations, many model sensitivity tests have better correlations than radar/model comparisons, indicating robustness in model performance on this aspect. It is further shown that well-validated model simulations could be used to constrain uncertainties in observed echo-top heights when the low-resolution surveillance scanning strategy is used.

1. Introduction

Full understanding of cloud and precipitation processes, and their impact on the surrounding environment, requires detailed information of precipitation structure and morphology, e.g., cloud-top heights and their spatial and temporal variations, the horizontal extent and the degree of organization of precipitation systems. For example, radar echo-top heights and their spatial and temporal distributions indicate the location of moistening and heating [e.g., *Masunaga et al.*, 2006; *Lau and Wu*, 2010; *Ahmed et al.*, 2016]; the spatial extent and organization of precipitation features are linked to different vertical distributions of latent heating, with small, scattered precipitation producing bottom-heavy heating, and large, organized precipitation systems resulting in top-heavy heating in a widespread stratiform region. Such detailed three-dimensional (3-D) information of cloud and precipitation structure comes from only two sources: radar observations and cloud-resolving model (CRM) simulations.

The DYNAMics of the Madden-Julian Oscillation (DYNAMO) field experiment (October 2011 – March 2012)¹ was carried out in the central tropical Indian Ocean to study MJO convective initiation processes and to improve MJO predictions [*Yoneyama et al.*, 2013]. Two sounding arrays together with multiple ground-based radars (Fig. 1) provide a rich dataset for studying 3-D precipitation structures and their associated environment during the MJO onset and mature stage. The current study focuses on the Northern Sounding Array (NSA) for the DYNAMO November MJO event because signals in temporal variations of this MJO event within the NSA are much stronger than in the Southern Sounding Array [*Johnson and Ciesielski*, 2013; *Johnson et al.*, 2015]. The NSA is defined by four sounding sites at Malé, Colombo, Addu Atoll, and the R/V Roger Revelle [*Ciesielski et al.*, 2014]. Multi-frequency, dual-polarization, ground-based radars were deployed both at the Addu site and onboard the R/V Roger Revelle. In addition, the space-borne Precipitation Radar (PR) onboard the Tropical Rainfall Measurement Mission (TRMM) satellite also provided 3-D echo structure over a larger domain than the ground-based radars, but with lower sensitivity and less frequent sampling. These radar observations have been compared in previous studies. For example, *Xu et al.* [2015] compared radar observed echo-top heights,

¹ DYNAMO is a joint project with CINDY2011 (Cooperative Indian Ocean Experiment on Intraseasonal Variability in the Year 2011), AMIE (ARM MJO Investigation Experiment), and LASP (Littoral Air-Sea Process).

stratiform rain fractions, and precipitation feature sizes at Addu Atoll, R/V Roger Revelle, and R/V Mirai. They concluded that despite some subtle differences in the rate of convective deepening, convection at Addu Atoll (0.7°S, 73.1°E) and R/V Roger Revelle (0°, 80.5°E), which were about 825 km apart along the equator, showed similar temporal variability throughout the MJO lifecycle, especially when compared to the R/V Mirai (8°S, 80.5°E) radar, located to the south of the equator, which showed nearly opposite precipitation patterns compared to the Revelle radar. *Xu and Rutledge* [2014], *Powell and Houze Jr.* [2015], and *Xu and Rutledge* [2015] compared ground-based radars and TRMM PR long-term observations and noted that the variability observed by the S-PolKa radar at Addu Atoll and the Revelle C-band radar were consistent with TRMM-PR observations. These radar observations therefore provide a useful, consistent database to validate CRM simulations, which, in turn, allows for an additional investigation into the commonalities and subtle differences between the different regions during the same MJO event.

CRMs are important tools for studying precipitation structure because of their ability to explicitly resolve deep convective dynamics, as opposed to parameterizing them in coarse resolution models. Generally, there are two different approaches for the limited area cloud-resolving simulations (usually with horizontal resolution of less than 4 km). One is to use CRMs in the forecast mode, where time-variant large-scale influences come from lateral boundaries. The other is to use the observed large-scale forcing and apply it uniformly in the model domain with cyclic lateral boundary conditions. Both approaches have been used in previous DYNAMO case studies. With the first approach, *Hagos et al.* [2014a, b] used the Weather Research and Forecasting (WRF) model to simulate DYNAMO MJO events and tested different microphysical packages. Although WRF captured many general features of the MJO evolution, it overestimated radar reflectivity for deep convective cells compared with the S-PolKa observations. *Wang et al.* [2015a] applied a regional setup of WRF and were also able to simulate the MJO characteristics as well as its moist static energy budget with a coarser horizontal spacing of 9 km. Other studies have used the second approach with cyclic boundary conditions and large-scale forcing. For example, *Janiga and Zhang* [2016] used the System for Atmospheric Modeling (SAM) [*Khairoutdinov and Randall*, 2003] to analyze the moisture budget during DYNAMO; *Wang et al.* [2015b] studied the cloud-radiation interactions using WRF. They identified significant

spread in radiative fluxes and stratiform cloud coverage when different microphysical schemes are used. Wang et al. [2016] further demonstrated important roles of horizontal advection of moisture and radiative feedback in driving the MJO in their WRF simulations with parameterized large-scale dynamics. These studies used the large-scale forcing approach because it guarantees accurate simulations of domain-mean moisture budget and surface precipitation. This provides a better constraint when the evolution of precipitation structures is validated against radar observations. Simulations using the first approach do not guarantee agreements in surface rainfall and its temporal variations. The current study uses the large-scale forcing approach.

The combination of observations from the multiple DYNAMO radars and different CRM simulations provides a unique opportunity to study the evolution of 3-D cloud and precipitation structures during convective initiation of the MJO, and their interactions with large-scale environment. The main challenge exists in interpreting discrepancies in simulations resulting from using different models, configurations, forcing, and observational validations. This is crucial for both CRM sensitivity studies and radar data interpretation aiming at understanding MJO mechanisms. The current study attempts to meet this challenge with three goals.

The first goal is to find out if, and to what extent, a CRM is able to simulate subtle differences in precipitation evolution of the same MJO event observed by radars at different locations with different sampling sizes. Observations by the S-PolKa radar on Addu Atoll [Rowe and Houze, 2014], the C-band radar onboard R/V Roger Revelle [Xu and Rutledge, 2015], and the TRMM PR covering the whole NSA area, including both Addu and Revelle sites, are analyzed to show similarities and differences in their convective properties and evolution. Their similarities are used to validate the Goddard Cumulus Ensemble (GCE) model [e.g., Tao et al., 2014], which is then used to understand the nature of the subtle differences in different radar observations.

The second goal of the current study is to quantify differences in simulated cloud structure and their variability during an MJO event resulting from using different CRMs and different model configurations. In this study, three CRMs, namely, GCE, WRF and SAM, are used with the identical large-scale forcing for the DYNAMO November MJO event. This approach almost

always results in desired matches in domain mean precipitation and total heating between the model and the observations. Once the mean state is realistically simulated, comparisons between the models and radar observations, as well as among models themselves, can be carried out with high confidence [e.g., *Fridlind et al.*, 2012; *Wang et al.*, 2015b].

The third goal of our study is to cross-compare simulations by different CRMs to observations from different DYNAMO radars. Previous model validation studies generally treat observations as the truth with pre-determined instrument errors. Multiple radar observations during the DYNAMO field campaign, together with radar-specific large-scale forcing for driving CRMs, may allow us to differentiate natural variabilities and observation uncertainties in precipitation structures. If the variability among radar observations is smaller than that among models, quantitative error margins may be derived for different model simulations. Otherwise, it could mean that the models perform within the error margin of radar observations and may be used, in some occasions, to help interpret differences in radar data. To achieve the third goal, several methods are used to reduce inter-model discrepancies. For example, all model simulations are carefully matched and coordinated. A radar simulator is used to avoid retrieval errors, and to match each radar's sampling strategy. Qualitative variables that are less sensitive to inter-model/inter-instrument differences (e.g., the 0-dBZ echo-top height) are used instead of more quantitative statistics such as radar Contour Frequency with Altitude Diagrams [*Yuter and Houze*, 1995].

The remainder of this study is divided into six sections: Section 2 describes DYNAMO sounding and radar data, as well as the three CRMs. Section 3 discusses GCE simulations using three different large-scale forcing datasets and compares them with corresponding radar observations in order to understand the natural variability of precipitation structures due to different radar sampling strategies. Section 4 compares simulations by three different CRMs using the same large-scale forcing. Section 5 presents sensitivity tests of the GCE model to its domain size and vertical resolution, as well as selected parameters that could affect the simulated echo-top heights of deep convection. Section 6 explores cross-validations using multi-radar and multi-model results and attempts to quantify both model and radar variability. Discussions and a summary are given in Section 7.

2. Radar and Model Descriptions

2.1 Radar Observations

During the DYNAMO campaign, two scanning radars were deployed on Addu Atoll: the dual-frequency S- (10 cm wavelength) and Ka- (0.86 cm wavelength) S-PolKa radar operated by the National Center of Atmospheric Research, and the C-band (5 cm wavelength) SMART-R radar operated by Texas A&M University. Both radars have excellent temporal coverage, but the S-band data from S-PolKa is used in this study due to its better spatial coverage compared with SMART-R, which experienced partial beam blockage, and also due to its additional Range Height Indicator (RHI) scans with improved vertical resolution. Gridded S-PolKa radar reflectivity, its retrieved surface rainfall rate, and convective/stratiform separation were provided by the University of Washington radar group. The S-PolKa radar has a scanning radius of 150 km and temporal resolution of 15 minutes. The surveillance (SUR) 360° azimuth scanning data, with only 8 elevation angles up to 11°, were interpolated onto 0.5 km Cartesian grids both horizontally and vertically. To compensate for the coarse vertical resolution of the SUR data, S-PolKa also performs the RHI scans at 1° azimuthal resolution up to 45° elevation, but for a limited scanning range toward the open ocean [Rowe and Houze, 2014]. In this study, SUR data were used for deriving the site-specific large-scale forcing, and will be used mainly for model validations. The RHI data were used to corroborate and validate SUR observations.

The R/V Roger Revelle was deployed due east of Addu Atoll at the equator (cf. Fig. 1). It carried out four separate cruises to the site during the campaign. As a result, there are gaps in radar observations, although most of the missing data periods are during the suppressed MJO phases. The NASA TOGA C-band (5-cm wavelength) radar data are processed and provided by the Colorado State University radar group. It has the same scanning radius of 150 km as the S-PolKa, but with a wider beam width (1.5° vs. 0.91° for S-PolKa). The data are gridded with 2-km horizontal resolution, 0.5-km vertical resolution, and 10-minute temporal resolution. Details of the C-band radar can be found in [Xu and Rutledge, 2014]. For the purpose of comparing radar observations with CRMs, only data points within the range between 20 km and 100-km range are

used in this study due to sampling errors. The rationale of using this range will be presented in section 6.

Both ground-based and space-borne radars provide highly valuable information on precipitation structures, with different characteristics in spatial and temporal coverage. Ground-based radars operate continuously and provide excellent temporal resolution. However, both the S-PolKa and C-band radar are at fixed locations and their spatial coverage are limited by their detection ranges. The space-borne radar, e.g., the Ku-band (2-cm wavelength) TRMM PR scanning radar, has large spatial coverage with a swath of about 250 km. However, its temporal coverage is relatively poor. Sampling frequency for the area covered by the NSA (Fig. 1) is about twice a day. TRMM PR's footprint is about 5 km, which is much coarser than the resolutions of the surface radars. However, it has a high vertical resolution at 250m. Its minimum detectable signal is about 17 dBZ, much less sensitive than the ground-based radars (better than 0 dBZ). The TRMM orbital data product 2A25 (version 7), post-processed to form the TRMM precipitation feature database [Liu *et al.*, 2008], is used to compare radar echo-top heights and precipitation feature sizes. TRMM surface rainfall data are based on gridded TRMM 3B42 product due to its high temporal resolution ($0.25^\circ \times 0.25^\circ$, 3-hourly) [Huffman *et al.*, 2007].

2.2 Large-Scale Forcing

Three independent radar observations with different sampling sizes during the DYNAMO field campaign provide a unique opportunity for model validations and comparisons. These radar observations are well correlated [Xu *et al.*, 2015] but still have differences due to their different locations and sampling methods. They are used in this study to test the CRMs' capability of reproducing both the similarities and differences in observations. In order to do this, three large-scale forcing datasets derived by two research groups independently using different methods are used to drive CRMs.

The NSA forcing was derived using objective budget analyses [Ciesielski *et al.*, 2014; Johnson *et al.*, 2015]. It uses 4 sounding stations defining the NSA (Fig. 1) to derive water vapor and momentum tendencies. Surface rainfall was derived from the water vapor budget. The temporal

resolution of the NSA forcing is 4 times daily. The NSA forcing used in this study is version 2a, which includes observations only without any global analysis data. The advantage of the NSA forcing is that it has large spatial coverage (roughly $7^{\circ} \times 7^{\circ}$) and is representative of the mean budget characteristics during the DYNAMO November MJO event. TRMM PR observations are used to compare with simulations driven by the NSA forcing over this large domain.

Large-scale forcing for both the Addu site (S-PolKa radar) and Revelle site (C-band radar) were derived by the Lawrence Livermore National Lab, using variational analysis [Xie *et al.*, 2010; Zhang *et al.*, 2001; Zhang and Lin, 1997]. The variational analysis uses domain-average surface and top-of-the-atmosphere radiation, surface fluxes and precipitation as constraints to adjust atmospheric state variables from sounding profiles by the smallest possible amount to conserve column-integrated mass, moisture, and static energy. ECWMF analysis data were used to derive the large-scale forcing 8 times daily. Surface precipitation amount uses radar-retrieved surface rainfall from the SUR scans. The matches between the large-scale forcing domain and the radar sampling domain ensures an apple-to-apple comparison between model simulations and radar observations.

2.3 Model Descriptions

Three different CRMs are used to simulate the November MJO event: the Goddard Cumulus Ensemble (GCE) model [e.g., Tao *et al.*, 2009], the System for Atmospheric Modeling (SAM) [e.g., Khairoutdinov and Randall, 2003; Janiga and Zhang, 2016], and the Weather Research and Forecasting (WRF) model version 3.5 [Skamarock *et al.*, 2008] modified for cloud-resolving simulations with prescribed or parameterized large-scale forcing [e.g., Wang and Sobel, 2011; Wang *et al.*, 2013; Anber *et al.*, 2014; Wang *et al.*, 2015b]. These simulations, referred to as inter-comparison simulations, have been carefully designed and coordinated to minimize uncertainties in their experimental setup. For example, all three models are driven by the same large-scale forcing (NSA), which includes advection of moisture and temperature, as well as domain mean horizontal winds. All three models use the same domain size, grid spacing, and start at the same time. Other numerical parameters, such as damping layers near the top of the model domain, as well as the associated damping coefficients, are also specified to be identical.

The inter-comparison simulations have $256 \times 256 \times 106$ grids, with horizontal grid spacing of 1 km. The nominal vertical grid spacing is 0.25 km, with a higher resolution in the planetary boundary layer, and a stretching, lower resolution from 18 km to about 30 km. All three models use the 2-moment Morrison microphysical scheme [Morrison *et al.*, 2005], although with slightly different versions inherited from every model's implementation. On the other hand, the CRMs differ in many other aspects. Each CRM has its own dynamics, numerical methods, and subgrid turbulent mixing parameterizations. Radiation schemes also differ (Table 1).

In addition to inter-comparison simulations, the GCE model is used for sensitivity studies on different forcing data, different microphysical packages, as well as varying domain size and vertical spacing. Table 1 summarizes some of the key parameters of CRM simulations. In order to differentiate ground-based radar data from their corresponding large-scale forcing, we will use the radar frequency (S-PolKa and C-band) when we refer to radar observations in our discussion, and use the location (Addu and Revelle) to represent model simulations using the radar specific large-scale forcing.

3. Comparisons of GCE Simulations and Radar Observations

Two different large-scale forcing data are derived separately for the Addu and Revelle site in order to account for the differences in their rainfall observations and associated environment. Because surface radar has limited coverage (~ 150 km radius), and the organized precipitation systems associated with an MJO can be larger than that, using large-scale forcing specific to a single radar and constrained by its own retrieved surface rainfall can significantly reduce uncertainties in radar/model comparisons. Furthermore, to ensure an apple-to-apple comparison, a forward radar simulator [Matsui *et al.*, 2009; Matsui *et al.*, 2013] has been used to derive radar reflectivity from model-simulated hydrometeors with the same microphysics assumptions. The more realistic Mie scattering is used for both the C- and Ku-band radar. S-band radar reflectivity is calculated using Rayleigh approximation, which is a reasonable assumption at longer wavelengths. GCE simulations using the large-scale forcing specific to each individual radar. They have a domain size of 256 km x 256 km, with the horizontal spacing of 1 km, and stretching vertical grids with the nominal resolution of 0.5 km. The microphysics used is the

Goddard 3-ICE package, which is a single-moment scheme including cloud, rain and three ice-phase species: ice, snow aggregates and graupel [Tao et al., 2009].

The CRM-simulated precipitation structures are compared with radar observations in both spatial and time dimensions. In the vertical, the 0-dBZ echo-top height distributions and their temporal variations are compared. In the horizontal dimension, the convective/stratiform separation and the precipitation feature sizes are compared. The precipitation feature size will also be used as an indication of convection organization. The lack of large precipitation features generally indicates scattered, less organized convection, and vice versa. Correlations between horizontal feature sizes and the maximum echo-top heights within the contiguous precipitation feature are also compared. Additional comparisons of basic model fields, e.g., the apparent heat source and moist sink, temperature profiles, are carried out against observations. They generally have very good agreements, as should be for a balanced large-scale forcing (plots not shown).

3.1 Surface Rainfall

Figure 2a and 2b show the S-PolKa radar and C-band radar retrieved mean surface rainfall with the corresponding GCE model simulations. Note that Addu forcing is only available starting November 13. For each independent forcing data, GCE model reproduces the mean surface rainfall trends very well. This is because the domain-mean heat and water vapor budget are constrained tightly by the prescribed large-scale forcing. The two different observations, however, agree with each other only qualitatively in that they all captured the enhanced surface rainfall during the mature stage of the MJO event (21-30 November), and the suppressed rainfall before and after. Both observations show two distinct rainfall peaks at around 24 and 27 November, indicating passages of two Kelvin wave events [e.g., Depasquale et al., 2014]. On the other hand, observations differ in quantitative details. For example, the C-band radar retrieved MJO rainfall peak lags S-PolKa retrieval and the NSA forcing derived rainfall peak by about 20 hours. This lag is because the Revelle was stationed about 7° to the east of Addu radar site [Xu et al., 2015]. The rainfall rate is more episodic at the Addu site, dominated by passages of individual events separated by low rainfall periods. For example, there are two consecutive high rain rate events during the developing stage around November 18. Furthermore, the peak

rainfall rate at Addu is twice as much compared with the C-band radar, indicating stronger convective events both during the developing and mature stages of MJO. Given the differences in surface rainfall observed at different locations, and ground-based radar's limited spatial coverage, it would be optimal to use location specific large-scale forcing, as in this study, to ensure accurate model and observation comparisons. For example, if one used the NSA forcing and compared the surface rainfall with the S-PolKa radar retrieval, the correlation would have been much lower. With the GCE model accurately reproducing mean surface rainfall variability, the next question to address is: How well can the CRMs quantitatively reproduce the variability of cloud structures during the MJO event, given the tight constraint of the surface rainfall? Whether the GCE model can reproduce subtle differences at different radar sites will also be a unique test for model performance.

3.2 Vertical Structure

Figure 3 compares vertical distributions of the 0-dBZ radar echo-top heights between the S-PolKa radar observations (3a) and GCE simulation using Addu forcing (3b); C-band radar observations (3c) and GCE simulation using Revelle forcing (3d). The 0-dBZ echo-top height is calculated by searching from the ground up for contiguous reflectivity values above 0 dBZ. All model data have been interpolated on the observation vertical levels. The black lines are the trends of median echo-top height using a 24-hour running mean. Both radars (Fig. 3a and 3c) show clear transition from a mixed cloud regime with low clouds dominant (10 to 22 November), to one dominated by deep convection (22 to 29 November). The onset of deep convection is earlier for S-PolKa (~22 November) than for C-band radar (~23 November), consistent with surface rainfall trends. In addition, temporal variations are more gradual in C-band radar observation, similar to [Xu *et al.*, 2015], where low clouds are shown to dominate between 10 and 17 November. The echo-top height distributions widen between 18 and 23 November, leading to mostly deep convection after November 23. On the other hand, S-PolKa observations show low-cloud dominating before and after the deep convection period (22 to 29 November). These general trends and subtle differences in observations are well captured in Figs. 3b and 3d, using the site specific large-scale forcing. This indicates the importance of using well-constrained, radar-specific large-scale forcing for quantitative model validations.

Several discrepancies exist in radar/model comparisons in Fig. 3. First, radar observed echo-top height distributions have larger variability compared with model simulations. For example, S-PolKa observed MJO mature phase between 23 and 29 November shows co-existing shallow precipitation (0-dBZ echo-top height between 3km and 5km) with dominant deep precipitation. *Barnes and Houze* [2013] also found similar co-existing shallow and deep echo during the MJO mature phase using TRMM PR data. The model simulation in Fig. 3b, on the other hand, shows very little shallow precipitation. Another discrepancy is that the GCE model misses the strongest convection compared with radar observations. This is more prominent in the S-PolKa radar comparison. For example, GCE simulations rarely produce any convection with 0-dBZ echo-top higher than 15 km (Fig. 3b), which are frequently observed by S-PolKa. This discrepancy becomes more obvious toward the end of the simulation. Another interesting feature that shows prominently in the simulation in Fig. 3d is the diurnal cycle during the transition from the suppressed phase to the developing phase (11 to 18 November). The tick marks on the x axis represent noon time in UTC. During this period, which corresponds to early morning local time, Fig. 3d shows high frequencies ($> 15\%$) of low clouds with echo-top heights below the 0°C level. Late afternoon sees a much wider cloud distribution and more cumulus congestus and deep convection, with only less than 5% shallow convection. C-band radar observations in Fig. 3c also show a similar diurnal cycle, but with less amplitude. Similarly, *Ruppert and Johnson* [2015] has analyzed the S-PolKa radar data and showed diurnal cycles in convective echo-top frequency and area between 13 and 16 November, as shown in Fig. 3a. Model simulation using the Addu forcing, on the other hand, does not show any diurnal cycle during the same period. One possible reason could be that the model is still in its spin-up period. Quantitative details of model simulated diurnal cycles during DYNAMO will be a subject of a future study.

3.3 Horizontal Structure

In terms of horizontal coverage, it is useful to categorize precipitation into convective and stratiform regions. The convective and stratiform region have distinct vertical structures, dynamics, and diabatic heating profiles [e.g., *Houze*, 2004]. Their partitioning can also be used as an important criterion for model validations. Figure 4 shows example scenes of observed and

modeled precipitation structures during the November MJO passage over the Addu site. The red area represents convective regions and the green the stratiform. An identical convective/stratiform separation algorithm based on *Steiner et al.* [1995] was applied to both observations and simulations to ensure accurate comparisons [Powell and Houze, 2013; Rowe and Houze, 2014]. Three snapshots are shown in Fig. 4: on November 18, during the developing phase and on November 23 and 24, during the MJO mature phase. Figure 4 illustrates the transition from isolated convection during the MJO developing phase, to organized convection during the initial period of MJO mature phase, to widespread stratiform rain at its later stage. On November 18, convection was scattered, with a small stratiform region surrounding each convective cell. The GCE model simulation can reproduce the overall structure. The November 23 time frame in the middle panel has more organized convection, with two large convective areas dominating the scene, and widespread stratiform region trailing them (the systems are moving to the east in general). The GCE simulation shows a well-organized convective line along with a wide stratiform region. However, the convective line consists of some smaller features compared with the larger contiguous convective region observed by S-PolKa. Ten hours later, the system became almost all stratiform, which is also well simulated by the model.

Figure 5 shows temporal variations of domain mean convective and stratiform rainfall rates from radar measurements (5a and 5c) and model simulations (5b and 5d), in order to provide a more quantitative assessment. Both the S-PolKa and C-band radar use the same convective/stratiform separation algorithm [Steiner et al., 1995], but with slightly different thresholds in S-PolKa [Rowe and Houze, 2014] and C-band [Xu and Rutledge, 2014]. Model simulated convective/stratiform separation matches corresponding radar algorithms. Figure 5 shows a good comparison at the Addu site, where the S-PolKa radar observed about 67% of total rainfall from convection, while the GCE simulated 68% convective rainfall. The area covered by convective rain is 20% in S-PolKa radar observations, compared with 26% in the model (plot not shown). In other words, the simulated stratiform area is less extensive in the simulation, but the rainfall rate agrees with the observation. In terms of temporal evolution, both observations and simulations show that stratiform rain tends to follow rises and falls of convective rain, but with several hours' delay in phase [e.g., Zuluaga et al., 2013]. This is because the stratiform precipitation is the result of organized deep convection. The mesoscale flows that support the widespread stratiform

region develop after convection peaks and organizes. At the Revelle site, Fig. 5d produces much more convective rainfall (89% of the total rainfall) compared with the observation (72%). The areal coverage, on the other hand, has better agreement (72% convective coverage observed by the C-band radar vs. 73% simulated). These results seem to indicate that in terms of convective/stratiform separation, the GCE model compares better with radar in more intense and organized convection observed at the Addu site. Interestingly, previous CRM simulations have shown similar results. For example, *Varble et al.* [2011] compared nine CRM simulations of a monsoonal event at Darwin, Australia, during the TWP-ICE field campaign. Similar to our Revelle case, they found that all models underestimated stratiform rainfall rate, despite the fact that some of the models actually overestimated the area of stratiform region. *Lang et al.* [2007] compared convective/stratiform separation during the TRMM Large-Scale Biosphere-Atmosphere (LBA) experiment in the Amazon in 1999 using different microphysical packages in the GCE model. Their comparisons with ground radar showed that the stronger, better-organized convection during the easterly wind regime compared better with the observation than the more oceanic, westerly regime case. These previous results are consistent with our DYNAMO case study. For example, the Addu forcing produces stronger, more organized convection, resulting in better model/radar comparisons, similar to the easterly flow case in *Lang et al.* [2007]. On the other hand, Revelle forcing produces more scattered, less organized convective systems, similar to the westerly flow case in *Lang et al.* [2007] and the *Varble et al.* [2011] monsoonal case, where the models overestimated convective rain. In conclusion, the GCE model can reproduce subtle differences in convection organization at the Addu and Revelle sites. On the other hand, it also inherited limitations and discrepancies in validating with less organized convection using convective/stratiform separations algorithms, as shown in several previous studies.

The sizes of contiguous precipitating area are correlated with rainfall intensity [*Ruppert Jr. and Johnson*, 2015] and is used as an indicator of the degree of convection organization [*Rowe and Houze*, 2014; *Xu et al.* 2015]. *Xu et al.* [2015] analyzed precipitation feature sizes measured by surface radars during the DYNAMO campaign by defining a precipitation feature as the region with contiguous radar reflectivity greater than 15 and 20 dBZ, respectively. Our study follows this method, but uses 17 dBZ as the threshold to be consistent with the minimum TRMM PR reflectivity. The effective size shown in Fig. 6 is defined as the square root of the total area of

each feature. Time series of the effective feature size distributions are shown in Fig. 6a for S-PolKa observations and Fig. 6b for GCE simulations using Addu forcing. Fig. 6c and 6d are for the C-band observations and model simulations using Revelle forcing. In all four panels, precipitation feature sizes increase with increasing echo-top heights and surface rainfall rates, a trend also described for all three DYNAMO MJO events by *Rowe and Houze* [2015]. The total numbers of features (shown in red dots in Fig. 6) tend to increase with feature size, too, although increases in feature numbers usually occur ahead of increases of feature sizes. This indicates that deep convection tends to start individually and later organizes themselves into larger convective systems, as shown also in Fig.5 and described by *Rowe and Houze* [2015]. Very large feature sizes appear with high frequency during the MJO mature phase, again indicating enhanced convection organization. Comparing S-PolKa (6a) and C-band radar (6c) observations during the MJO mature phase (21 to 30 November), we find that S-PolKa (6a) observed higher frequencies of both the very large feature size (more than 40 km) and very small feature size (less than 5 km) compared with C-band radar observations (6c). The high frequencies of the very large features are consistent with the previous conclusion that the Addu site has more organized convective systems and more stratiform rain. GCE simulations using both Addu forcing (Fig. 6b) and Revelle forcing (Fig. 6d) can reproduce temporal variations of the observed convective feature sizes. However, both simulations overestimate feature sizes. The largest discrepancy occurs during the mature stage when the mean feature size peaks. This shows that the model has over-organized convection within the domain. Previous studies have shown that the low-level wind shear is crucial to organizing convections [e.g., *LeMone et al.*, 1998]. Since the large-scale forcing provides only domain mean wind field, it may have discouraged random convection from developing in these simulations. In addition to using 17 dBZ as the threshold for precipitation features, we have tested an alternative algorithm where the convective/stratiform separation was used. For example, the contiguous convective region was defined as a “convective feature”. This later method resulted in smaller feature sizes, but all conclusions remain the same as shown in Fig. 6.

After deriving both the vertical extents and horizontal sizes of precipitation features, the natural next step is now to examine their relationship. For every precipitation feature identified, we tally its effective length and the maximum radar echo-top height. Fig. 7 shows contours of the joint

probability density function for S-PolKa radar observations (7a), GCE simulations using Addu forcing (7b), C-band radar observations onboard Reville (7c), and GCE simulations using Reville forcing (7d). Large feature size is a sufficient condition for high radar echo-tops, but it is not necessary in all four cases. This is because large-size features generally consist of many individual convective elements at various development stages, with at least some of them being deep. On the other hand, many relatively small, isolated deep convection do reach the tropopause. Among the four panels in Fig. 7, the S-PolKa observations have by far the largest amount of precipitation features with their echo-top heights above 15 km, extending to 20 km, while the simulations in Fig. 7b and 7d have few. The C-band observations also shows some precipitation features above 15km, but with less frequency compared with S-PolKa. The model simulated feature size vs. maximum echo-top height are shown in Figs. 7b and 7d. For small features with effective length less than 30 km, simulated maximum echo-top height distributions agree reasonably well with observations, especially at the Reville site. However, the Addu site simulations are missing all features with echo-top height above 15 km. The biggest discrepancy exists for large precipitation features, e.g., features with effective lengths longer than 70 km. For example, for an S-PolKa observed feature with 100 km effective length, the echo-top heights range between 13km and 20km, whereas the GCE simulated height ranges are between 10km and 15km. Similar discrepancies also exists at the Reville site. Given the fact that the simulated mean surface rainfall agrees well with the observations, and that previous studies show CRMs overestimating updraft strengths in deep convection with the TWP-ICE case study [Varble *et al.*, 2014; Wu *et al.*, 2009], we speculate that the underestimation of echo-top heights in CRMs may be due to the model deficiency in ice-phase microphysical processes or numerical damping issues. Sensitivity tests are carried out in section 5 to test these hypotheses.

3.4 Comparisons with TRMM Precipitation Radar

Ground-based radar has the advantage of continuous coverage with a high spatial resolution. However, it has limited spatial coverage. The space-borne radar, e.g., the TRMM PR precipitation radar and CloudSat radar, has a global coverage, but with limited sampling frequencies at a fixed location. For example, the NSA area shown in Fig. 1 has only 2 to 3 TRMM overpasses per day, far less than the 10- to 15-minute interval of ground-based scanning

radar. Space-borne radar also has lower sensitivity compared with the ground-based radar. During DYNAMO period, CloudSat radar did not collect much data due to a battery malfunction. In this section, we compare GCE model simulation with the TRMM PR data using a similar methodology as in sections 3.2 and 3.3.

To ensure a fair comparison, we performed the third set of simulations using NSA large-scale forcing derived from the four sounding stations in the NSA. The TRMM PR scanning data are used if at least one contiguous data point is located within the NSA box. In other words, the TRMM sampling domain is larger than the NSA box, although at least part of the precipitation feature must locate within the NSA. Figure 8 compares 17-dBZ radar echo-top height distributions (upper panel) and the same variable simulated by the GCE model. Again, the TRMM PR radar simulator is used to derive radar reflectivity from the GCE model simulation. The low temporal resolution in TRMM PR is obvious in Fig. 8a. Also, due to TRMM PR's low sensitivity, the 17-dBZ echo-top height distributions are contaminated by the large amount of data points cluttered near the radar bright band caused by melting ice-phase particles. It is difficult to use TRMM PR data quantitatively for this single case study, as shown in Fig. 8, because of the high noise level. However, the observations and model agree qualitatively in 17-dBZ echo-top height distributions. The convection is dominated by low clouds with 17 dB echo-top height below the melting level during the suppressed phase of MJO (10 to 15 November). At the mature phase (24 to 28 November), these low clouds occur significantly less. GCE simulated TRMM PR data in Fig. 8b still use the hourly data output from the model. It shows that the 17-dBZ echo-top height almost completely missing below 5 km between 24 and 28 November, indicating exclusive deep convection and widespread deep stratiform rain. The temporal variations of the median echo-top height (black lines in Fig. 8) do show variations around the melting level, but the amplitudes are much smaller than surface-based radar observations in Fig. 3.

The simulated precipitation feature sizes are also compared with TRMM PR data. Due to limited temporal resolution of TRMM PR, all features are aggregated over the 30-day simulation period to show the probability distributions of the precipitation feature sizes in Fig. 9. The black line is the TRMM observation and the red line is the GCE simulation. Precipitation features with sizes

less than 50 km², which correspond to about two PR sampling points, are not considered. In general, the distributions agree very well between the observations and simulations, both of which peak at around 80 km² size and drop off quickly toward large-size features. The simulation slightly overestimates medium features around 200 km² size and underestimate features larger than 1000km². When the feature size becomes large, the probability of the feature intercepting with the edge of either the TRMM PR swath or model lateral boundaries increases. The sizes of these edge features are simply counted as they are. This means that the feature size distribution is biased at the large tail of the size spectrum and only represents the distributions within the sampling swath. However, the comparison remains meaningful because the model domain (256 km) is comparable to the TRMM swath width (250 km).

Similar to Fig. 7, correlations between 17-dBZ precipitation feature size and its maximum echo-top heights for each feature are plotted in Fig. 10. It generally confirms the conclusion drawn from Fig. 7. Good comparisons between the TRMM PR observations and the GCE simulation exist for features with effective lengths less than 50 km. The simulation again misses the strongest radar echo-top heights for very large features.

4. Model Inter-comparison

Three different CRMs: GCE, SAM and WRF, are used in model inter-comparisons in this section. The NSA large-scale forcing is used for these simulations in order to represent precipitation variations over the large area covered by the four sounding stations. The simulations have been coordinated to use the same specifications as much as possible to reduce uncertainties in results (Table 1). The matching radar observation for NSA forcing is the TRMM PR data, as shown in section 3.4. However, TRMM data suffer from low sensitivity of 17 dBZ. In order to avoid losing information at less than 17 dBZ reflectivity regime, we focus on presenting only model results in this section. Figure 2c shows surface rainfall comparisons among three CRMs (colored lines), together with forcing derived surface rainfall (the black line). Different model simulations agree with one another very well. The GCE model using the Morrison scheme is not plotted in Fig. 2c because it is almost identical to the GCE 3-ICE scheme simulation. When compared with forcing-derived surface rainfall, models generally agree better

with observations when rainfall rates are high and the forcing is strong, but miss some of the peaks during the suppressed and transition period. This could be because of the large uncertainties when the large-scale forcing is weak. *Johnson et al.* [2015] have compared the forcing-derived surface rainfall with TRMM 3B42 product, as well as several other global surface rainfall products, and found good agreements among the observations during the DYNAMO period. The rest of this section will again focus on precipitation structure comparisons. Since we are only comparing model results in this section, the Rayleigh assumption is used to calculate radar reflectivity for its simplicity.

Figure 11 shows time series of 0-dBZ echo-top height distributions for the four sets of simulations. The GCE model uses both Goddard 3-ICE scheme (Fig. 11a) and Morrison two-moment scheme (Fig. 11b). The SAM (Fig. 11c) and WRF (Fig. 11d) model simulations both use the Morrison scheme. In general, temporal variations of echo-top height distributions agree well among all simulations. The median echo-top heights, shown as black lines with 24-hour running mean, correlate well with one another. They demonstrate a clear transition from low clouds dominating the suppressed stage (10 to 15 November), to the more diverse and evenly distributed cloud heights during the developing stage (16 to 21 November), where low, middle and high clouds coexist. Deep clouds dominate the mature stage (22 to 27 November) with the shallow clouds population almost completely missing, and large areas of deep stratiform cloud deck present. This shows the robustness of different CRMs in simulating precipitation structures when they are constrained by the same observed large-scale forcing. Compared with simulations using radar site specific large-scale forcing in Fig. 3, NSA forcing produces more robust and gradual, less episodic distributions, probably because the NSA domain is about five times the size of individual radar domains. Averaging over the large domain results in smoother temporal variations. This again indicates the importance of matching large-scale forcing location with observation domain for model validations. On the other hand, comparing simulations using NSA forcing and using Addu/Revelle forcing also indicates that the former is more desirable in terms of reproducing robust MJO phase change signals. Subtle differences still exist among different models. For example, the small fraction of convection that reaches above 12 km is largely missing in Fig. 11b, but appears with more frequency for other simulations. In general, GCE model simulates less frequencies of these deep convection compared with both the SAM and

WRF model. On the other hand, the SAM model (Fig. 11c) has lower echo-top heights during the suppressed phase, and the transition from the low clouds to mixed clouds seems to be more abrupt, compared with the other two models.

Precipitation feature sizes and their temporal variations are shown in Fig. 12. The black lines represent median feature sizes, and the red lines represent the total feature number. Figure 12 again shows consistent temporal variations in terms of feature sizes. Compared with the episodic structures in Fig. 6, NSA forcing produces a clearer transition from the suppressed to active phase in the horizontal structures. During the suppressed phase and at the beginning of the developing phase (10 to 17 November), all precipitation feature sizes remain less than 20 km, indicating small, scattered convection. The echo-top height distributions in Fig. 11, on the other hand, show that starting from November 15, deep convection has already started to develop. The precipitation features remain small and scattered for an additional day or two before starting to increase in sizes on November 17. This is an indication that the convection starts to organize only after 1~2 day of the onset of medium to deep convection. Features larger than 60 km in sizes mainly appear during the mature phase (22 to 27 November), representing well-organized convective systems. The largest precipitation features occurred between 24 and 25 November, when occasionally the whole domain consists of one or two contiguous features. They were labeled as missing data in the plots, as shown by the gaps around that period. Despite general similarities among different CRMs, GCE produces larger sizes and fewer precipitation features compared with both SAM and WRF simulations during the suppressed phase. On the other hand, the WRF model has the smallest feature sizes and the highest total feature number during the suppressed and developing stage between 12 and 21 November.

5. Sensitivity Tests

All model simulations discussed so far used the domain size of 256 km x 256 km. One thing we have not addressed is matching model domain sizes with the area size where the large-scale forcing is derived. While the 256 x 256 km domain matches the single-radar sampling diameter (~300 km), it is much smaller than the NSA sounding array size of 830 km x 770 km. The small domain used in section 4 for NSA forcing may have limitations in representing some of the

mesoscale convective systems, especially during the mature stage of the MJO, when some of the observed mesoscale convective systems have physical sizes larger than the 256 km domain size (cf. Fig. 4). This could produce low biases in simulated feature sizes since compensating updrafts and downdrafts must be balanced within the limited domain, possibly limiting the largest of feature sizes. Small model domain size and a coarse vertical resolution both could affect echo-top height distributions and their temporal variations with MJO phases. In order to address these uncertainties, we conduct additional sensitivity tests using GCE model with the Goddard 3-ICE microphysical scheme and the NSA forcing. The small domain simulation uses the same 256×256 km size as in section 4, but with a coarser vertical resolution of 45 levels, instead of 106 levels as used in model inter-comparisons. The large-domain simulation has the dimension of 1024×1024 , also with 45 vertical levels. The horizontal resolution remains the same as 1 km. The vertical levels are also stretched, with the nominal resolution of 0.5 km at the middle troposphere.

Figure 13a and 13b show distributions of the 0-dBZ echo-top heights for the small- (Fig. 13a) and large- (Fig. 13b) domain simulation. The median echo-top heights, shown as the black lines, again correlate with each other well, with clearly defined suppressed, developing and mature stages. These results suggest that the GCE simulation with 0.5-km vertical resolution can produce similar general cloud distribution characteristics as the higher resolution of 0.25 km. However, it does so with more noise and less confidence compared with both higher vertical resolution and larger domain size simulations. The increased sample size in the large domain simulation (Fig. 13b) produces much smoother temporal variations in echo-top height distributions compared with Fig. 13a. For example, during the suppressed phase on November 12, there is a strong deep convective event that dominates in Fig. 13a, resulting in a spike in the median echo-top height. The big domain simulation, on the other hand, can still maintain enough shallow clouds so that the median echo-top height remains low. When comparing Fig. 13a with the small-domain, fine vertical resolution simulation in Fig. 11, the later again shows less noise. The local maximum on November 12 is less prominent in all panels in Fig. 11 compared with Fig. 13a. This suggests that insufficient vertical resolution could lead to spurious strong convection under certain circumstances, possibly because of biases in parameterized turbulent mixing and its sensitivity to grid sizes.

Another set of sensitivity tests are designed to explore possible causes of models' underestimation of echo-top height in deep convection (e.g., Fig. 3). Note that these sensitivity tests may not be physically realistic, but serve the purpose of isolating and elucidating possible physical mechanisms that may affect echo-top heights. Three possible mechanisms have been identified: reducing graupel sizes, removing spurious cloud evaporation, and raising damping layers at the top boundary. In deep convection, graupel is the dominant hydrometeor species contributing to radar reflectivity at upper levels. In a bulk microphysical scheme, graupel size distributions are assumed to be either exponential or gamma. These assumptions are empirical and often have large uncertainties. The working hypothesis is that, reducing graupel's mean sizes reduces their terminal fall velocities accordingly. As a result, graupel particles could be lifted to higher levels by the same updrafts, increasing simulated radar echo-top heights. To test this hypothesis, the intercept parameter for the assumed exponential graupel size distribution in the Goddard 3ICE scheme was doubled in the first sensitivity test. The resulted echo-top height distributions are shown in Fig. 13c. Compared with Fig. 13a, Fig. 13c has very little change. This test suggests that graupel size uncertainties in microphysical scheme are not likely reasons for the underestimation of deep echo-top heights.

The previous sensitivity test shows that reducing hydrometeors' terminal fall velocity does not help in increasing cloud echo-top heights in the current case study. A natural speculation follows is that the strengths of the updraft cores in deep convection might be too weak. Cloud-scale vertical motion are not observed during DYNAMO, but we can design sensitivity tests with a CRM to increase simulated updraft velocities. Two experiments, one changes microphysics, the other changes the numeric, are carried out in this section. The first experiment reduces the so-called spurious evaporation, which is produced by artificial effect of cloud boundaries advecting through a Eulerian grid where excessive evaporation may be produced [Klaassen and Clark, 1985; Grabowski and Morrison, 2008; Lang et al., 2014]. The cooling generated by the spurious evaporation could act to damp deep convection. In the GCE sensitivity test, a condition has been added that the cloud droplets/ice does not evaporate/sublime, even in unsaturated environment, when the vertical air velocity is higher than 0.5 m/s. The reduces the evaporation at cloud edges where cloudy scene occupies only part of the numerical grid [Lang et al., 2014]. The resulted

echo-top heights are shown in Fig. 13d. Reducing spurious evaporation indeed resulted in increased echo-top heights, by an average value of about 2 km. However, there are still little convection that reached above 15 km. On the other hand, reduced evaporation also invigorated shallow convection, especially during the suppressed period. The high frequency of shallow convection with echo-top heights between 1 and 4 km on November 4, and between November 10 and 15 are significantly reduced, resulting in unrealistic temporal variations of echo-top heights compared with radar observations. In conclusion, although artificially removing spurious cloud evaporation produced taller deep convection, it also invigorates shallow convection during the suppressed MJO phase which is unrealistic. Therefore this is not the main reason for the CRM deficiency. The next experiment in the attempt to increase vertical velocity involves raising the numerical damping layer near the upper boundary from 18 km to 22 km. This damping layer is needed in CRMs to absorb vertically propagating gravity waves. However, excessive damping at upper levels sometimes results in weakened updrafts, especially at higher levels. The model results, as shown in Fig. 13e, are again very similar to the control study shown in Fig. 13a, indicating that upper level damping is not the reason for underestimation of echo-top heights.

To summarize, all simulations, with different forcing, different models and microphysics, as well as different vertical resolutions and domain sizes, have reproduced observed cloud population characteristics and the timing of transitions from the suppressed, to developing, to the mature phase of the November MJO event. However, they show different noises and offer different confidence level when interpreting the results. For CRM simulations using large-scale forcing, it is recommended that the domain size should in general match the area covered by the sounding array, in order to better carry out the model/observation comparisons. Higher vertical resolution is also desirable in reducing the noise level in echo-top height variations and providing more details in its distributions. Three sensitivity tests designed to increase radar echo-top heights in deep convection failed to identify reason(s) for the underestimation of cloud echo-top heights. Instead, these results suggest that uncertainties in cloud microphysics, as well as upper level numerical damping, are unlikely to be the underlying mechanisms. Future study will focus on investigating convection-circulation feedbacks in the Tropics and their impacts on limited-area models driven by large-scale forcing.

In terms of horizontal extent of simulated precipitation features, the large-domain simulation has added a small amount of larger than 256 km precipitation features. Figure 14 shows the scatter plot of the 17-dBZ effective feature lengths vs. the maximum echo-top heights for every precipitation feature. Figure 14a and 14c are for the two small-domain simulations with 0.5km (Fig. 14a) and 0.25km (Fig. 14c) vertical resolution, respectively. The largest feature size in these two panels is roughly the same as their domain size of 200 km. The main sensitivity to the vertical resolution is that higher vertical resolution in Fig. 14c can simulate shallow precipitation with high rainfall rates, as shown by the points below 4 km. These points are missing with the 45-level simulation in Fig. 14a. However, this does not necessarily mean that there is no shallow precipitation in the coarser vertical resolution simulation; they just did not make the threshold of 17 dBZ (figure not shown). In other words, the coarser resolution simulation produces weaker shallow convection compared with its finer vertical resolution counterpart. When the 1024 km x 1024 km domain is used, maximum precipitation feature size increases to 370 km (Fig. 14b), comparable to the largest size observed by TRMM PR (Fig. 14d). As in all GCE simulations using NSA forcing (and to a lesser extent for both SAM and WRF simulations), the deepest convection with echo-top height above 15 km is again largely missing. The large features have their maximum 17-dB echo-top heights centered at around 12 km, whereas TRMM PR shows them scattered between 12 km and 19 km. This confirms the previous conclusion that CRM simulations underestimate echo-top heights, especially for large, contiguous convective systems. Future studies are needed to resolve this issue.

6. Model and Radar Cross Validations

Comparisons between multiple model simulations and radar observations have demonstrated that the three CRMs can robustly reproduce structural characteristics of precipitation at different MJO phases and their temporal evolution. Furthermore, subtle differences shown at different radar sites for the same MJO event are well captured by models when constrained by the site-specific large-scale forcing. After achieving positive results in model validations, we now focus on the third goal of this study: to quantify inter-model and inter-radar discrepancies in

precipitation structures in order to guide better interpretations of model simulations and radar observations.

Two variables are used to quantify vertical distributions of radar echo-top heights: correlation coefficients of the median echo-top heights, representing their temporal variations as the MJO progresses, and differences between the mean of the time series. Figure 15 summarizes scatter plots of these two variables for various pairs of observations/simulations. All correlation coefficients are calculated with strictly matched time stamps. A high correlation coefficient combined with a small mean difference suggest a good agreement. In Fig. 15, circles represent model-model comparisons; squares represent model-observation comparisons; and the triangles are radar-radar comparisons. Simulations constrained by the same large-scale forcing are shown in warm-colored circles, whereas purple circles represent GCE simulations using different large-scale forcing. The label for each point indicates the single parameter that differs. For example, the purple circle labeled “NSA/Revelle” represents the comparison between two GCE model simulations with identical setups otherwise, but one using NSA forcing, the other using Revelle forcing. Model sensitivity tests are clustered in two groups: the warm-colored ones at the high correlation/small difference corner, and the purple ones on the low correlation side. This shows that using the specific large-scale forcing derived for the observation site is of the first order importance for accurately representing temporal variations of echo-top heights. Among the warm-colored circles, red circles compare three different models, all using the NSA forcing. SAM and WRF agree better on the mean echo-top heights, whereas GCE and WRF agree better in terms of temporal variations. The main cause of differences in echo-top heights is that the GCE model has higher echo tops compared with the other two during the suppressed phase between 10 to 15 November (cf. Fig. 11). For shallow convection, two factors might contribute to the differences: surface fluxes and the sub-grid mixing schemes (cf. Table 1). Additional studies are needed to clarify the causes. Figure 15 also shows sensitivities using the GCE model. Switching microphysical schemes from Goddard 3-ICE to Morrison scheme has a small effect on simulated precipitation structures, with the correlation coefficient above 0.9 and the mean echo-top height difference of 0.1 km. Increasing GCE’s vertical resolution mainly affects echo-top heights (mean differences of 0.7 km), but not on their temporal variations (> 0.9 correlation coefficient). On the other hand, when model domain size increases, the main impact is on

temporal correlations. This information can potentially guide future modeling studies. For example, modelers often face choices when they cannot have both large-domain and high-resolution at the same time due to computational constraints. Figure 15 indicates that increasing vertical resolution should be chosen if accurate simulations of the cloud vertical structure is the goal. However, in order to achieve robust temporal variations, one should first consider using a large enough domain.

The squares in Fig. 15 show comparisons between model simulations and radar observations. Blue squares are comparisons between C-band radar observations and the Revelle forcing simulations; green squares are between S-PolKa observations and the Addu forcing simulation. S-PolKa radar observations and Addu forcing simulation have slightly better correlation at 0.53, compared with 0.38 for the C-band observations and Revelle forcing simulation. The fact that they are all lower than correlation coefficients among model sensitivity tests indicates that, for our future sensitivity studies, the choices of model/microphysical scheme/domain/vertical resolution likely will not affect results pertaining to temporal variations of echo-top heights, at least to the extent that can be meaningfully verified by surface radar observations.

The consistency in model performances encourages us to try to corroborate radar observations with model simulations. The challenge here is that the S-PolKa radar-observed 0-dBZ echo-top heights have a dependency on data range due to the coarse vertical resolution when using the SUR scanning mode. This is because radar beam widens with distances. When reflectivity data were re-gridded, larger uncertainties exist at locations farther away from the radar site, especially for limited elevation angles during the S-PolKa SUR scanning. Three different ranges for radar observations are shown in Fig. 15: the full dataset with the range between 0 and 150 km, “range 1” with data located between 20 km and 120 km, and “range 2” with data between 20 km and 100 km. The reason for using 20 km as the minimum range is because the radar cannot detect the actual echo top of deep convection within the close range. There are two sets of radar/radar comparisons: S-PolKa SUR (SPol) vs. C-band Revelle radar (C-band) at different locations, and S-PolKa’s two different scanning strategies (SPol vs. RHI in Fig. 15). For the C-band observations, the average echo-top height decreases slightly from 6.13 km for the full dataset, to 6.10 km for the 20~120 km range, to 6.05 km for the 20~100 km range. For the S-PolKa data,

they are much larger at 6.90 km, 6.47 km and 6.09 km, respectively. This is the main reason for the vertical spread of the green squares. Model simulations consistently show smaller differences than the S-PolKa-observed range of mean echo-top height. If we assume the model simulation is correct, we can deduce that data subset which has the best model-radar comparison is the most accurate representation of the true 3D reflectivity field. According to Fig. 15, this is “range 2” data, subset between 20 km and 100 km. This conclusion can be also verified independently using the multi-radar/multi-model strategy presented in this study, as well as by the same S-PolKa radar with its two different scanning methods, labeled as SPol/RHI in Fig. 15. Assuming model simulations were accurate, the simulated differences between Addu forcing and Revelle forcing should match the observed differences by S-PolKa and the independently measured C-band radar. This is exactly the case for “range 2” data shown in Fig. 15, in the pair of points labeled as “Addu/Revelle” and “SPol/C-band Range 2”. For SPol/RHI comparisons, the temporal correlations are all around 0.7, showing a high degree of consistency in the temporal variations measured by the two different sampling methods. They are higher than the squares in Fig. 15, which represent model and radar comparisons, with correlation coefficients between 0.35 and 0.5. In other words, model simulated temporal variations in echo-top heights can be further improved using radar observations. In terms of the mean echo-top height, the differences between SUR and RHI scanning range from 0.6 km for the full scan, to 0.2 km for 20~100km scan range 2. The least difference between SPolKa SUR scan and RHI scan (SPol/RHI) is for range 2 data, the same conclusion drawn when using radar/model comparisons. The decision of using range 2 subset for echo-top height comparisons for both S-PolKa and C-band radar is based on cross validation results presented in this section.

7. Summary

The unique aspect of the current study is its use of observations from three radars and three CRMs driven by radar site-specific large-scale forcing to study the temporal variations of precipitation structures, both vertically and horizontally, during the DYNAMO November MJO event. Observations from the S-PolKa radar at Addu Atoll, TOGA C-band radar onboard R/V Roger Revelle, and the space-borne TRMM PR, simulations by cloud-resolving models GCE, SAM and WRF, and three large-scale forcing derived for the two ground-based radar sites at

Addu and Revelle, and over the NSA, were compared within a consistent framework of radar data analysis, modeling, and model validations. Sensitivity of the GCE model to microphysical schemes and processes, domain sizes, and vertical resolutions were also tested.

The study emphasizes how precipitation structures change as an MJO event evolves, which is closely related to variations of diabatic heating and the moisture budget during the MJO event. Two direct radar measurables, the 0-dBZ echo-top height and the 17 dBZ precipitation feature size, were used to quantify precipitation structures. The largest uncertainty inherent in these measurables was radar's sampling sizes and scanning strategies, as discussed in section 6. Frequency distributions of the echo-top heights represented the intensity of convection, and the feature sizes indicated the degree of organization of convection. A radar simulator was used to calculate radar reflectivity for both the C- and Ku-band radar to ensure accurate comparisons. During the DYNAMO period, global MJO signals were characterized by the Real-time Multivariate MJO (RMM) indices [*Wheeler and Hendon, 2004*]. The current study categorized DYNAMO MJO phases in the Indian Ocean into four stages according to local cloud structures observed by radar: (1) The period from November 1 to 9, which corresponded to the global analysis of RMM phases 4, 5, and 6, represented the gap phase where the peak precipitation has moved out of the Indian Ocean. Precipitation was characterized by moderate rainfall with a mixture of low, medium and high clouds and a relatively small amount of organized convection. (2) November 10 to 16 was the suppressed phase (RMM phases 7 and 8) where the rainfall rate was low and convection was dominated by scattered shallow precipitation within the NSA. (3) During the MJO developing phase from November 17 to 23 (corresponding to RMM phase 1), surface rainfall and echo-top heights started to increase. Precipitation had a mixture of convection with different heights. Some convection also started to organize, forming larger precipitation features. (4) The mature phase spanned from 24 to 30 November (corresponding to RMM phases 1, 2 and 3). At the mature stage, deep clouds dominated and eventually organized into large contiguous features. At the later stage of the mature phase, stratiform cloud decks dominated the scene. These common characteristics were consistent with previous studies and showed robustly in all radar observations and model simulations during the DYNAMO November MJO event.

Another goal is to take advantage of the multi-radar, multi-model framework to identify subtle differences in precipitation structures at different locations: with S-PolKa and C-band radar near the equator, roughly 825 km apart at the two southern corners of the NSA, and TRMM PR covering the whole NSA domain. This provides a new opportunity for additional model validations and cross-comparisons. For example, compared with the C-band radar, the S-PolKa radar observed more deep convection and higher mean echo-top heights for the entire November MJO event. Convection at the S-PolKa site was also more episodic. During the MJO mature stage, higher frequencies of large precipitation features and higher percentages of stratiform coverage were observed at the S-PolKa site, compared with the C-band site. These subtle differences were all captured reasonably well by the GCE model when site-specific large-scale forcing was used, and when the matched convective/stratiform separation algorithms were used for both the simulations and observations. TRMM PR has limited temporal sampling within the NSA. However, model simulations using NSA forcing compared well within TRMM PR's detection threshold, both in terms of radar echo-top heights and distributions of precipitation feature sizes. On the other hand, certain systematic errors still existed for all model simulations. For example, the models missed the deepest convection, especially during the mature stage within large precipitation features. This needs to be investigated in the future.

The multi-radar, multi-model cross comparisons provided by the DYNAMO field campaign not only increased our confidence in both observed and simulated precipitation structures during the November MJO event, they also provided a unique opportunity to explore uncertainties in different radar scanning strategies and how they affected model/radar comparisons. As shown in this study, echo-top heights observed by the S-PolKa radar strongly depended on sampling ranges due to its limited sampling. Combining C-band observations and model simulations, we have identified S-PolKa subset data with a 20 km to 100 km sampling range as the best match, and subsequently used the results to help model validations.

Acknowledgment:

This study is mainly funded by Grant DE-SC0008568 of the US Department of Energy, Regional and Global Climate Modeling Program and Atmospheric System Research Program. X. Li would like to acknowledge additional funding from NASA grant NNX13AQ29G. S. Wang acknowledges support from National Science Foundation under grants AGS-1062206, AGS-1305788, and AGS-1543932. W.-K. Tao is also supported by the NASA Precipitation Measurement Mission (PMM) and NASA Modeling Analysis and Prediction (MAP) program. Both S-PolKa and C-band radar data are archived at DOE ASR's data site <https://asr.science.energy.gov/data>. TRMM PR data can be obtained through NASA website <https://pmm.nasa.gov/data-access/downloads/trmm>. The model simulations are archived on DOE server <http://portal.nersec.gov/project/cpmmjo> and NASA's mesoscale model webserver <https://cloud.gsfc.nasa.gov>. The RMM index is from www.bom.gov.au/climate/mjo/. Drs. Shaocheng Xie and Yunyan Zhang from Lawrence Livermore National Lab, and Dr. Paul Ciesielski from Colorado State University provided large-scale forcing data and many insights in the interpretations. We would also like to thank Dr. Samson Hagos at PNNL for his help in archiving data. Dr. Zhe Feng at PNNL and Prof. Courtney Shumacher at Texas A&M provided valuable suggestion on using surface-based radar data. Acknowledgment is also made to the NASA Center for Climate Simulation, NASA Advance Supercomputing Division, and NASA Precipitation Processing System, for resources used in this research. PMEL Contribution # 4684 (CZ).

Reference:

- Ahmed, F., C. Schumacher, Z. Feng, and S. Hagos (2016), A retrieval of tropical latent heating using the 3d structure of precipitation features, *J. App. Meteor. Clim.*, 55, 1965-1982.
- Anber, S., S. Wang, and A. H., Sobel (2014), Response of atmospheric convection to vertical wind shear: cloud resolving simulations with parameterized large-scale circulation. Part I: Specified radiative cooling. *J. Atmos. Sci.* 71, 2976-2993.
- Barnes, H. C., and R. A. Houze (2013), The precipitating cloud population of the Madden-Julian Oscillation over the Indian and west Pacific Oceans, *J. Geophys. Sci.*, 118, 6996-7023.

- Ciesielski, P. E., H. Yu, R. H. Johnson, K. Yoneyama, C. N. Long, J. Wang, S. M. Loehrer, K. Youngg, S. F. Williams (2014), Quality-controlled upper-air sounding dataset for DYNAMO/CINDY/AMIE: Development and corrections, *J. Atmos. Ocean. Tech.*, 31, 741-764.
- Depasquale, A., C. Schumacher, and A. Rapp (2014), Radar observations of MJO and Kelvin wave interactions during DYNAMO/CINDY2011/AMIE, *J. Geophys. Res.*, 119, 6347-6367, doi:10.1002/2013JD021031.
- Fridlind, A. M., A. S. Ackerman, J.-P. Chaboureaud, et al. (2012), A comparison of TWP-ICE observational data with cloud-resolving model results, *J. Geophys. Res.*, 117, D05204, doi:10.1029/2011JD016595.
- Grabowski, W. W., and H. Morrison (2008) Toward the mitigation of spurious cloud-edge supersaturation in cloud models. *Mon. Wea. Rev.*, 136, 1224–1234, doi:<https://doi.org/10.1175/2007MWR2283.1>.
- Hagos, S., Z. Feng, C. D. Burleyson, K.-S. Lim, C. N. Long, D. Wu, and G. Thompson (2014a), Evaluation of convection-permitting model simulations of cloud populations associated with the Madden-Julian Oscillation using data collected during the AMIE/DYNAMO field campaign, *J. Geophys. Res. Atmos.*, 119, 12,052-12,068, doi:10.1002/2014JD022143.
- Hagos, S., Z. Feng, K. Landu, and C. N. Long (2014b), Advection, moistening, and shallow-to-deep convection transitions during the initiation and propagation of Madden-Julian Oscillation, *J. Adv. Model Earth Sys.*, 6, 938-949, doi:10.1002/2014MS000335.
- Houze, R. A. (2004), Mesoscale convective systems, *Rev. Geophys.*, 42, RG4003, doi:10.1029/2004RG000150.
- Huffman, G. J., D. T. Bovin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, E. F. Stocker (2007), The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeor.*, 8, 38-55.
- Janiga, M. A., and C. Zhang (2016), MJO moisture budget during DYNAMO in a cloud-resolving model, *J. Atmos. Sci.*, 73, 2257-2278.
- Johnson, R. H., and P. E. Ciesielski (2013), Structure and properties of Madden—Julian Oscillations deduced from DYNAMO sounding arrays, *J. Atmos. Sci.*, 70, 3157-3179.
- Johnson, R. H., P. E. Ciesielski, J. H. Ruppert Jr., and M. Katsumata (2015), Sounding-based thermodynamic budgets for DYNAMO, *J. Atmos. Sci.*, 72, 598-622.
- Khairoutdinov, M. F., and D. A. Randall (2003), Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities, *J. Atmos. Sci.*, 60, 607-625.
- Klaassen, G. P., and T. L. Clark (1985), Dynamics of the cloud-environment interface and entrainment in small cumuli: Two-dimensional simulations in the absence of ambient shear, *J. Atmos. Sci.*, 42, 2621-2641.
- Lang, S., W.-K. Tao, R. Cifelli, W. Olson, J. Halverson, S. Rutledge, and J. Simpson (2007), Improving simulations of convective systems from TRMM LBA: Easterly and westerly regimes, *J. Atmos. Sci.*, 64, 1141-1164.
- Lang, S., W.-K. Tao, J.-D. Chern, D. Wu, X. Li (2014), Benefits of a 4th ice class in the simulated radar reflectivities of convective systems using a bulk microphysics scheme, *J. Atmos. Sci.*, 71, 3583-3612, doi:<http://dx.doi.org/10.1175/JAS>.

- LeMone, M. A., E. J. Zipser, and S. B. Trier (1998), The role of environmental shear and thermodynamic conditions in determining the structure and evolution of mesoscale convective systems during TOGA COARE, *J. Atmos. Sci.*, 55, 3493-3518.
- Liu, C., E. J. Zipser, D. J. Cecil, S. W. Nesbitt, and S. Sherwood (2008), A cloud and precipitation feature database from 9 years of TRMM observations, *J. Appl. Meteor.*, 47, 2712-2728 doi:10.1175/2008JAMC1890.1.
- Matsui, T., X. Zeng, W.-K. Tao, H. Masunaga, W. Olson, and S. Lang (2009), Evaluation of long-term cloud-resolving model simulations using satellite radiance observations and multi-frequency satellite simulators, *J. Atmos. Ocean. Tech.*, 26, 1261-1274.
- Matsui, T. et al. (2013), GPM satellite simulator over ground validation sites, *Bull. Amer. Meteor. Soc.*, 94, 1653-1660, doi: <http://dx.doi.org/10.1175/BAMS>.
- Morrison, H., J. A. Curry, and V. I. Khvorostyanov (2005), A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description, *J. Atmos. Sci.*, 62, 1665-1677.
- Powell, S. W., and R. A. Houze Jr. (2015), Evolution of precipitation and convective echo top heights observed by TRMM radar over the Indian Ocean during DYNAMO, *J. Geophys. Res. Atmos.*, 120, 3906-3919, doi:10.1002/2014JD022934.
- Rowe, A. K., and R. A. Houze Jr. (2014), Microphysical characteristics of MJO convection over the Indian Ocean during DYNAMO, *J. Geophys. Res.*, 119, 2543-2554, doi:10.1002/2013JD020799.
- Rowe, A. K., and R. A. Houze, Jr. (2015), Cloud organization and growth during the transition from suppressed to active MJO conditions. *J. Geophys. Res. Atmos.*, 120, 10,324–10,350, doi:10.1002/2014JD022948.
- Ruppert Jr, J. H., and R. H. Johnson (2015), Diurnally modulated cumulus moistening in the preonset stage of the Madden–Julian Oscillation during DYNAMO, *J. Atmos. Sci.*, 72, 1622-1647.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X. Huang, W. Wang, and J. G. Powers (2008), A description of the Advanced Research WRF version 3, NCAR Tech. Note NCAR/TN–475+STR, 125 pp.
- Steiner, M., R. A. Houze Jr., and S. E. Yuter (1995), Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data, *J. Appl. Meteor.*, 34, 1978-2007.
- Tao, W.-K. et al. (2009), A Goddard multi-scale modeling system with unified physics, *Annales Geophysicae*, 27, 3055-3064.
- Tao, W.-K. et al. (2014), The goddard cumulus ensemble (gce) model: Improvements and applications for studying precipitation processes., *Atmos. Res.*, 143, 392-424.
- Varble, A., A. M. Fridlind, E. J. Zipser, A. S. Ackerman, J. P. Chaboureau, J. Fan, A. Hill, S. A. McFarlane, J. P. Pinty, and B. Shipway (2011), Evaluation of cloud-resolving model intercomparison simulations using TWP-ICE observations: Precipitation and cloud structure, *J. Geophys. Res.*, 116, D12206, doi:10.1029/2010JD015180.
- Varble, A., E. J. Zipser, A. M. Fridlind, P. Zhu, A. S. Ackerman, J. P. Chaboureau, S. Collis, J. Fan, A. Hill, and B. Shipway (2014), Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations. Part I: Deep convective updraft properties, *J. Geophys. Res.*, 119, 13,891-13,918, doi:10.1002/2013JD021371.

- Wang, S., and A. H. Sobel (2011), Response of convection to relative sea surface temperature: Cloud resolving simulations in two and three dimensions, *J. Geophys. Res.*, 116, D11119.
- Wang, S., A. H. Sobel, and Z. Kuang (2013), Cloud-resolving simulation of TOGA-COARE using parameterized large-scale dynamics. *J. Geophys. Res.*, 118, doi:10.1002/jgrd.50510.
- Wang, S., A. H. Sobel, F. Zhang, Y. Q. Sun, Y. Yue, and L. Zhou (2015a), Regional simulation of the october and november mjo events observed during the CINDY/DYNAMO field campaign at gray zone resolution, *J. Clim.*, 28, 2097-2119.
- Wang, S., A. H. Sobel, A. Fridlind, Z. Feng, J. M. Comstock, P. Minnis, and M. L. Nordeen (2015b), Simulations of cloud-radiation interaction using large-scale forcing derived from the CINDY/DYNAMO northern sounding array, *J. Adv. Model Earth Sys.*, 7, 1472-1498, doi:10.1002/2015MS000461.
- Wang, S., A. H. Sobel, and J. Nie (2016), Modeling the mjo rain rates using parameterized large scale dynamics: Vertical structure, radiation, and horizontal advection of dry air, *J. Adv. Model. Earth Syst.*, 8, 121-139, doi:10.1002/2015MS000529.
- Wu, J., A. D. Del Genio, M.-S. Yao, and A. B. Wolf (2009), WRF and GISS SCM simulations of convective updraft properties during TWP-ICE, *J. Geophys. Res.*, 114, D04206, doi:10.1029/2008JD010851.
- Xie, S., T. Hume, C. Jakob, S. Klein, R. B. McCoy, and M. Zhang (2010), Observed large-scale structures and diabatic heating and drying profiles during TWP-ICE, *J. Clim.*, 23, 57-79.
- Xu, W., and S. A. Rutledge (2015), Morphology, intensity and rainfall production of MJO convection: Observations from DYNAMO shipborne radar and TRMM, *J. Atmos. Sci.*, 72, 623-640.
- Xu, W., and S. A. Rutledge (2014), Convective characteristics of the Madden-Julian Oscillation over the central Indian Ocean observed by shipborne radar during DYNAMO, *J. Atmos. Sci.*, 71, 2859-2877.
- Xu, W., S. A. Rutledge, C. Shumacher, and M. Katsumata (2015), Evolution, properties, and spatial variability of MJO convection near and off the equator during DYNAMO, *J. Atmos. Sci.*, 72, 4126-4147.
- Yoneyama, K., C. Zhang, and C. N. Long (2013), Tracking pulses of the Madden-Julian Oscillation, *Bull. Amer. Meteor. Soc.*, 94, 1871-1891.
- Yuter, S. E., and R. A. H. Jr. (1995), Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part II: Frequency distributions of vertical velocity, reflectivity, and differential reflectivity., *Mon. Wea. Rev.*, 123, 1941-1963.
- Zhang, M. H., and J. L. Lin (1997), Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to arm measurements, *J. Atmos. Sci.*, 54, 1503-1524.
- Zhang, M. H., J. L. Lin, R. T. Cederwall, J. J. Yio, and S. C. Xie (2001), Objective analysis of ARM IOP data: Method and sensitivity, *Mon. Wea. Rev.*, 129, 295-311.
- Zuluaga, M. D., and R. A. Houze, Jr. (2013), Evolution of the population of precipitating convective systems over the Equatorial Indian Ocean in Active Phases of the Madden-Julian Oscillation. *J. Atmos. Sci.*, 70, 2713-2725.

1109
1110
1111
1112
1113
1114
1115
1116
1117
1118

Table 1: Summary of CRM simulations' setup and model parameters.

Model	Equations	Microphysics	Radiation	Surface Fluxes	Subgrid mixing	Forcing	Domain (km)	$\Delta x, \Delta y, \Delta z$ (km)
SAM	anelastic	Morrison	RRTM	Calculated w/ SST	Smagorinsky	NSA	256×256×30	1, 1, 0.25
WRF	compressible	Morrison	RRTMG (shortwave) Goddard (longwave)	Calculated w/SST	Smagorinsky	NSA	256×256×30	1, 1, 0.25
GCE	anelastic	Morrison or Goddard 3ICE	Goddard	Prescribed w/ observation	Prognostic TKE	NSA	256×256×30 or 1024×1024×30	1, 1, 0.25 or 1, 1, 0.5
						Addu		
						Revelle		

Figure Captions:

Fig. 1: Schematic plot of DYNAMO sounding array and radar locations. The sounding observations from Malé, Colombo, Gan Island of Addu Atoll, and R/V Revelle are used to derive the NSA large-scale forcing. The yellow circles indicate sounding frequency of 4 times daily; red circles of 8 times daily. Radars were deployed on Addu Atoll, R/V Revelle and R/V Mirai. Adapted from *Ciesielski et al.* [2014].

Fig. 2: Surface rainfall observed (black lines) and simulated (colored lines) during the November MJO event for (a) the Addu forcing simulation, (b) Revelle forcing simulation, and (c) NSA forcing simulation.

Fig. 3: Comparisons of the 0-dBZ radar echo-top height distributions observed by S-PolKa radar (3a) and C-band radar (3c), and simulated by the GCE model using Addu forcing at the S-PolKa site (3b) and Revelle forcing at the C-band radar site (3d). The color images are frequencies of the 0-dBZ height occurrences. The black lines represent the median of the 0-dBZ echo-top heights with a 24-hour running mean.

Fig. 4: Comparisons of the convective/stratiform separation snapshots between S-PolKa observed (upper row) and GCE simulated (lower row) scenes at different MJO phases. Red represents convective area and green presents stratiform. The three columns illustrate transitions from isolated convection (left column at 00 UTC on Nov. 18) to organized convection (middle column at 14 UTC on Nov. 23) to deep stratiform clouds (right column at 00 UTC on Nov. 24) during the passage of the November MJO event at the Addu site.

Fig. 5: Time series of convective vs. stratiform rainfall rate observed by S-PolKa radar (5a) and C-band radar (5c), and simulated by GCE with Addu forcing (5b) and Revelle forcing (5d). The black lines are the convective rain rates; the red lines are the stratiform rain rates.

Fig. 6: Time series of the convective feature size frequency distributions (color shaded) observed by S-PolKa radar (6a) and C-band radar (6c), and simulated by the GCE using the Addu forcing (6b) and Revelle forcing (6d). Red lines represent the total numbers of features within the observation/simulation domain; black lines represent the mean sizes of the convective features, both of which are 24-hour running means.

Fig. 7: Contour plots of probability density function of the 17-dBZ feature length vs. the maximum 0-dBZ echo-top height for every individual precipitation feature. Fig. 7a is the S-PolKa radar observation. Fig. 7b is the GCE simulation using the Addu large-scale forcing. Fig. 7c is the C-band radar observation, and Fig. 7d is the GCE simulation using Revelle forcing. Both S-PolKa and C-band use only data within 100 km range.

Fig. 8: Comparisons of the 17-dBZ radar echo-top height distributions observed by the TRMM PR radar (8a) and simulated by the GCE model using NSA forcing (8b). The

color images are frequencies of the 17 dBZ height occurrences. The black lines represent the median echo-top heights with a 24-hour running mean.

Fig. 9: Comparisons of precipitation feature size distributions, defined by the 17-dBZ near surface radar reflectivity. The black line is the TRMM PR observation between November 1 and December 1, 2011, within the NSA. The red line is the GCE model simulated feature size distribution over the same period.

Fig. 10: Contours of probability density function of the 17-dBZ precipitation features' effective lengths vs. the maximum 17-dBZ echo-top heights. Fig. 10a is the TRMM PR observation over the NSA in November 2011. Fig. 7b is the GCE simulation using NSA forcing over the same period.

Fig. 11: The same as Fig. 3, except for CRM simulations using the NSA forcing for the month of November 2011. Fig. 11a is the GCE simulation using the Goddard 3ICE microphysical scheme; 11b is the GCE simulation using the Morrison two-moment scheme; 11c is SAM and 11d is WRF simulation, both using the Morrison two-moment scheme. The dark lines represent the median cloud-top heights with a 24-hour running mean.

Fig. 12: Same as Fig. 6, except for different CRM simulations. Fig. 12a is the GCE model with the 3ICE microphysical scheme; 12b is the GCE model with the Morrison two-moment microphysical scheme; 12c is the SAM model, also with Morrison scheme; 12d is the WRF model with Morrison scheme.

Fig. 13: The same as Fig. 11, except for GCE model with Goddard 3ICE scheme, using coarser vertical resolutions of 45 vertical levels instead of 106 as shown in Fig. 11. Fig. 13a is the small-domain simulation using 256km x 256km horizontal domain with 1km resolution, as in all previous simulations. Fig. 13b uses 1024 km x 1024 km horizontal domain with 1-km resolution. Fig. 13c is the same as Fig. 13a, except for reducing graupel sizes in the Goddard microphysical scheme. Fig. 13d is the same as Fig. 13a, except for removing cloud evaporation and ice sublimation where vertical air velocities are higher than 0.5 m/s. Fig. 13e is the same as Fig. 13a, except for raising the upper boundary sponge layer 3km higher. The dark lines again represent the median cloud top heights with a 24-hour running mean.

Fig. 14: Scatter plots of the precipitation feature effective sizes vs. the maximum 17-dBZ echo-top heights, showing feature sizes up to 400 km for large-domain simulation comparisons. Fig. 14a and 14c are the small-domain simulations using the 256x256 km domain size. Fig. 14a has 45 vertical levels and Fig. 14c has 106 vertical levels. Fig. 14b is the large-domain simulation with 1024x1024 km domain size and 45 vertical levels. Fig. 14d is the TRMM observation. Fig. 14d (14c) is the same as Fig. 10a (10b) except for an extended x-axis range.

Fig. 15: Summary of the radar echo-top height variations during the DYNAMO November MJO event, using various observations and simulations. The x-axis is the correlation coefficients of the median echo-top height time series (the black lines in

previous figures); the y-axis is their mean differences. Circles represent model-model comparisons; triangles are radar-radar comparisons; squares are model-radar comparisons. Each label indicates the single parameter that differs in the comparison pair, where the first label always has a higher echo-top height. In observations, range 1 represents sub-sampling of the original data range (0 km to 150 km), between 20 km and 120 km; range 2 is from 20 km to 100 km.

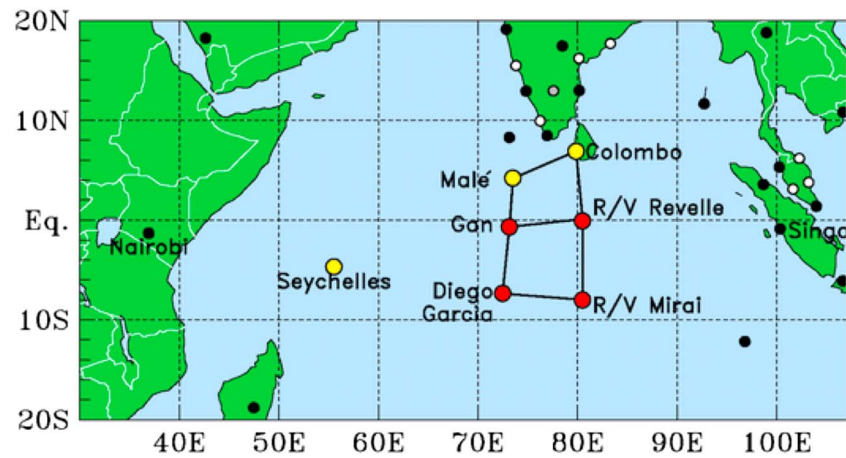


Figure 1. Schematic plot of DYNAMics of the Madden-Julian Oscillation sounding array and radar locations. The sounding observations from Malé, Colombo, Gan Island of Addu Atoll, and R/V *Revelle* are used to derive the NSA large-scale forcing. The yellow circles indicate sounding frequency of 4 times daily; red circles of 8 times daily. Radars were deployed on Addu Atoll, R/V *Revelle*, and R/V *Mirai*. Adapted from Ciesielski et al. (2014).

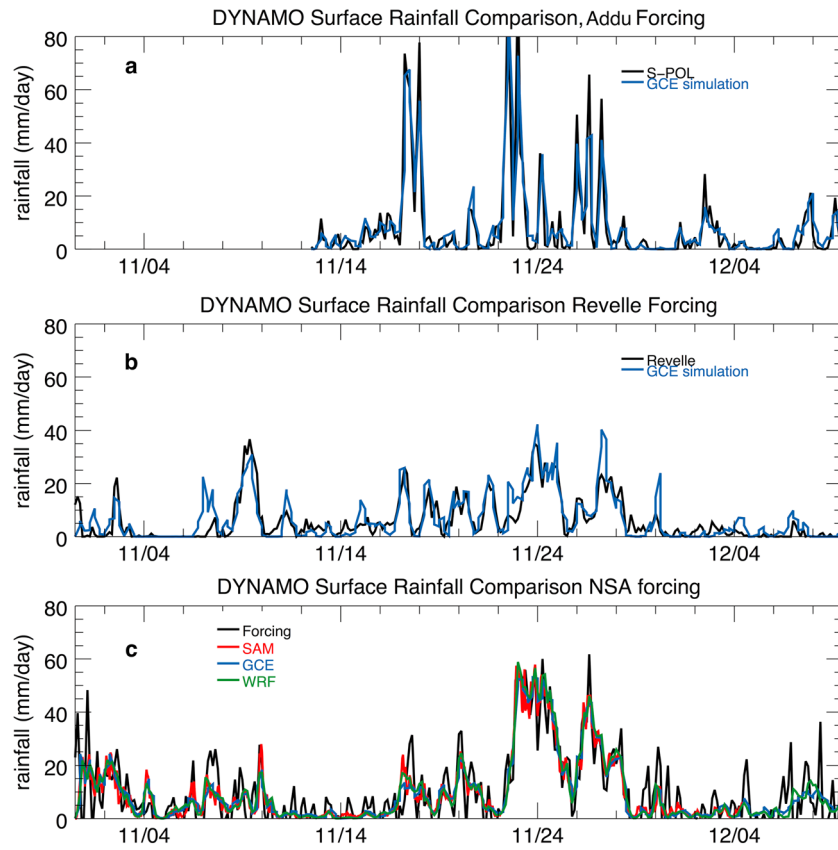


Figure 2. Surface rainfall observed (black lines) and simulated (colored lines) during the November Madden-Julian Oscillation event for (a) the Addu forcing simulation, (b) Revelle forcing simulation, and (c) northern sounding array (NSA) forcing simulation. DYNAMO: DYNAMics of the Madden-Julian Oscillation; GCE = Goddard Cumulus Ensemble; SAM = System for Atmospheric Modeling; WRF = Weather Research and Forecasting.

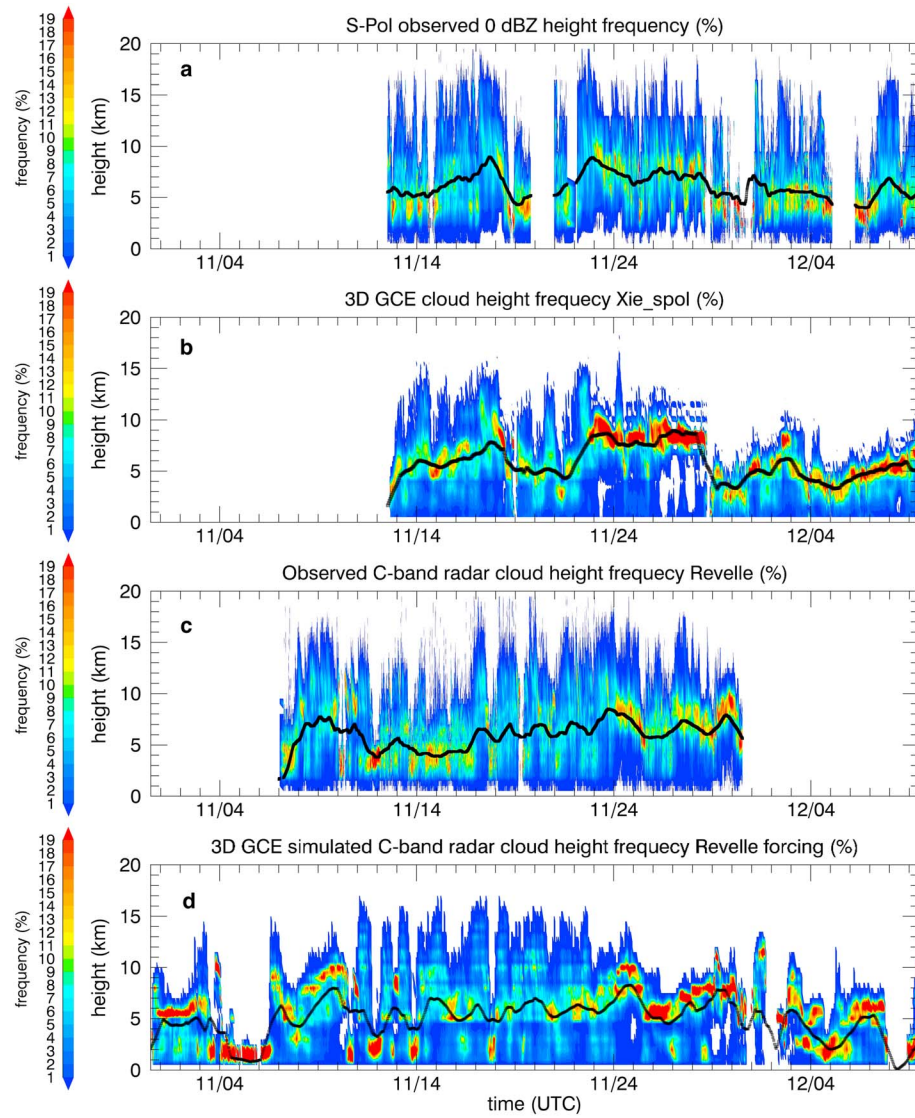


Figure 3. Comparisons of the 0-dBZ radar echo-top height distributions observed by S-PolKa radar (a) and C-band radar (c) and simulated by the GCE model using Addu forcing at the S-PolKa site (b) and Revella forcing at the C-band radar site (d). The color images are frequencies of the 0-dBZ height occurrences. The black lines represent the median of the 0-dBZ echo-top heights with a 24-hr running mean. GCE = Goddard Cumulus Ensemble.

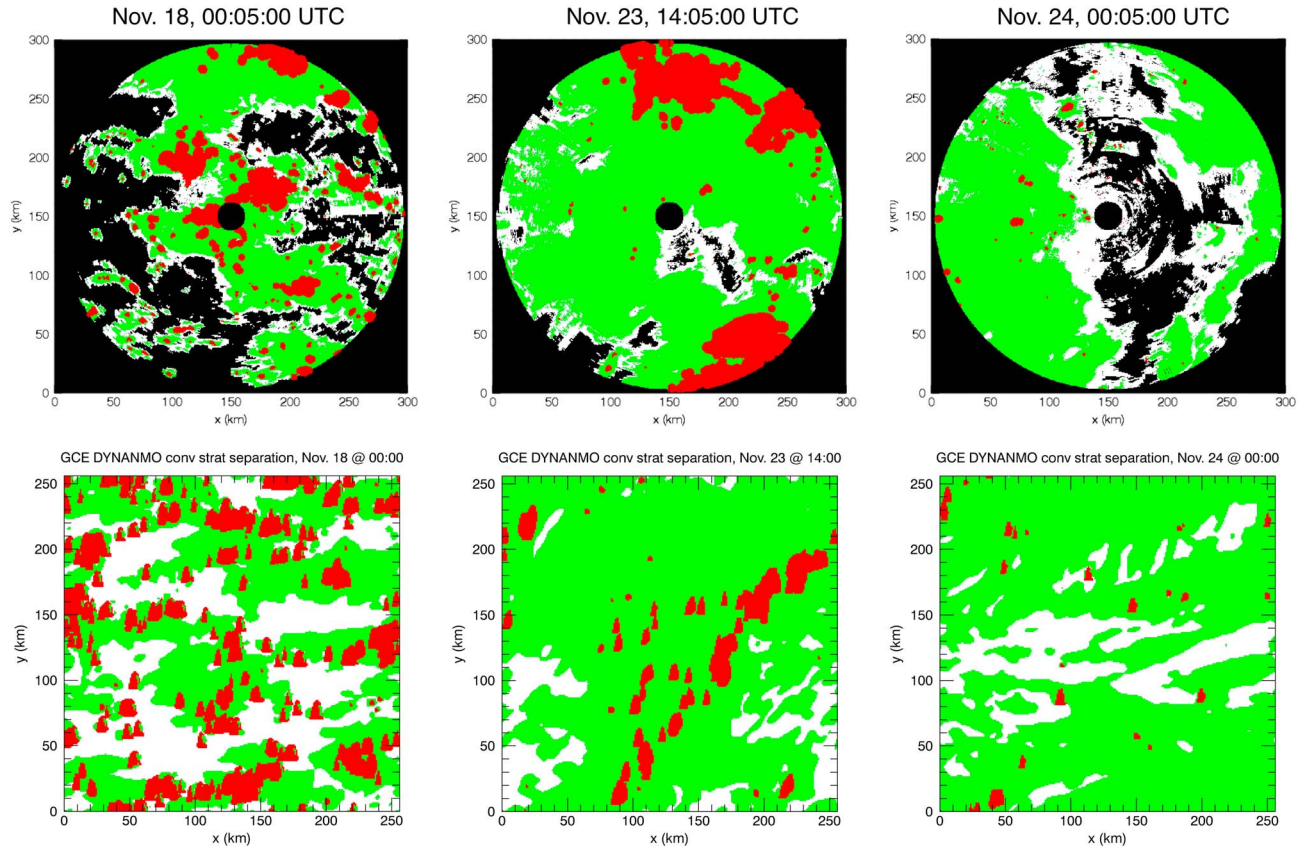


Figure 4. Comparisons of the convective/stratiform separation snapshots between S-PolKa observed (upper row) and GCE simulated (lower row) scenes at different MJO phases. Red represents convective area and green presents stratiform. The three columns illustrate transitions from isolated convection (left column at 00 UTC on 18 November) to organized convection (middle column at 14 UTC on 23 November) to deep stratiform clouds (right column at 00 UTC on 24 November) during the passage of the November Madden-Julian Oscillation event at the Addu site. GCE = Goddard Cumulus Ensemble; DYNAMO: DYNAMics of the Madden-Julian Oscillation.

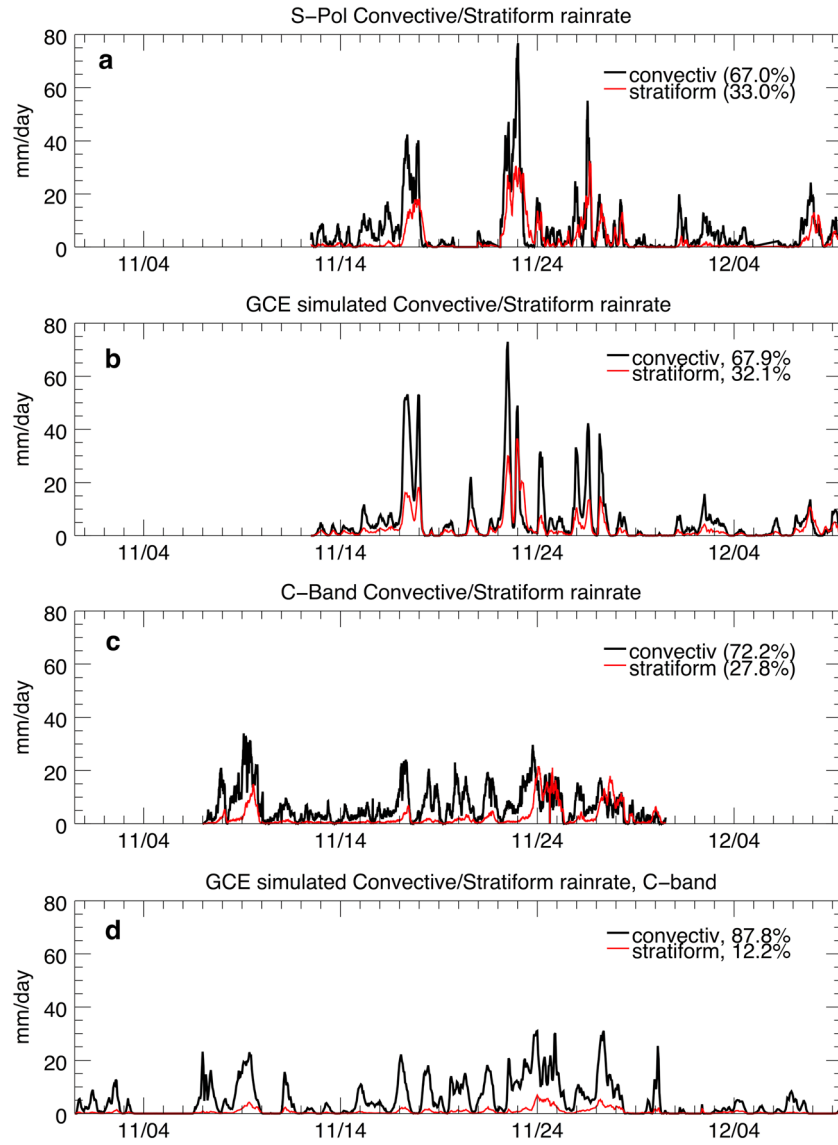


Figure 5. Time series of convective versus stratiform rainfall rate observed by S-PolKa radar (a) and C-band radar (c) and simulated by Goddard Cumulus Ensemble (GCE) with Addu forcing (b) and Reville forcing (d). The black lines are the convective rain rates; the red lines are the stratiform rain rates.

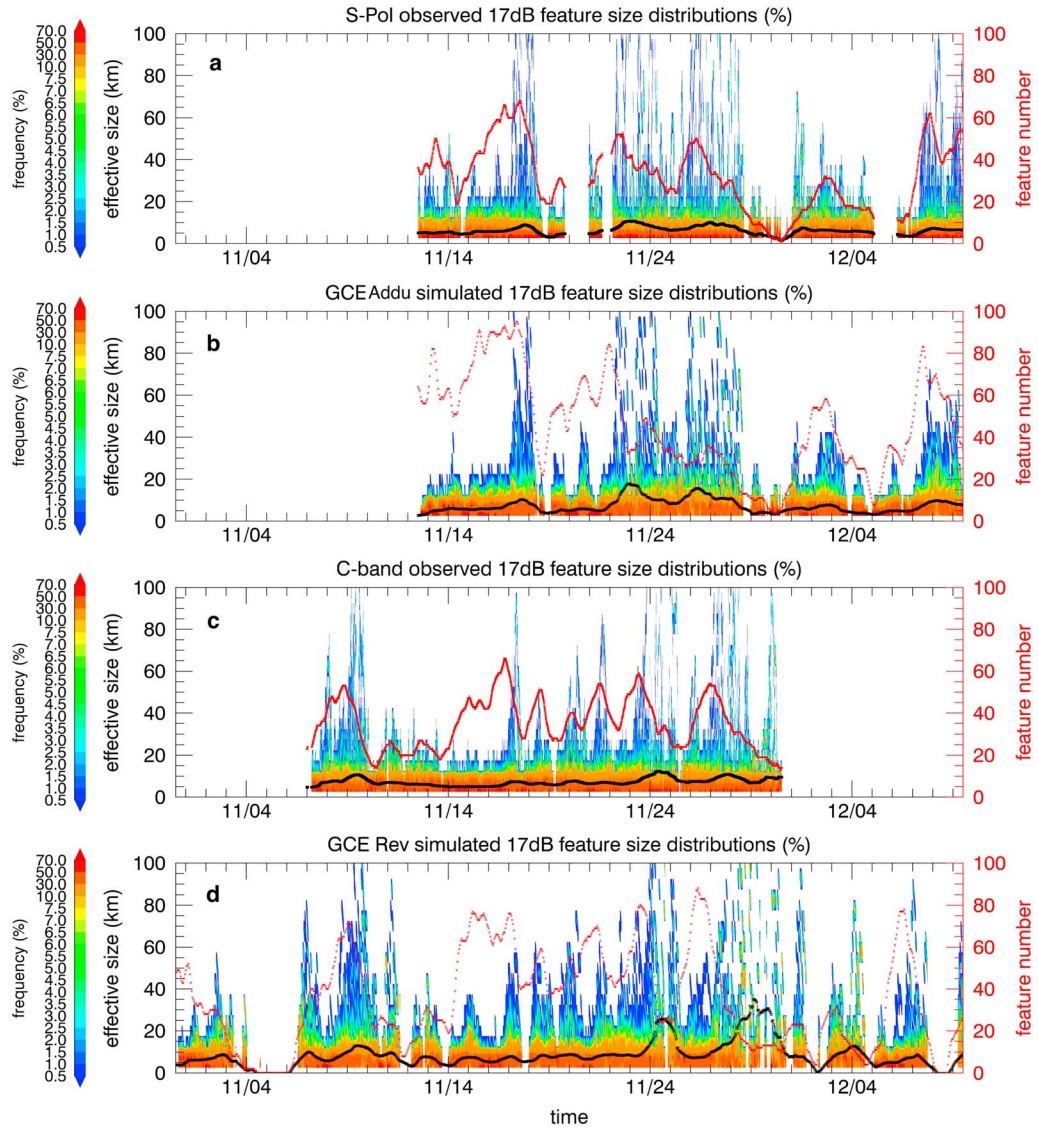


Figure 6. Time series of the convective feature size frequency distributions (color shaded) observed by S-PolKa radar (a) and C-band radar (c) and simulated by the Goddard Cumulus Ensemble (GCE) using the Addu forcing (b) and Revelle forcing (d). Red lines represent the total numbers of features within the observation/simulation domain; black lines represent the mean sizes of the convective features, both of which are 24-hr running means.

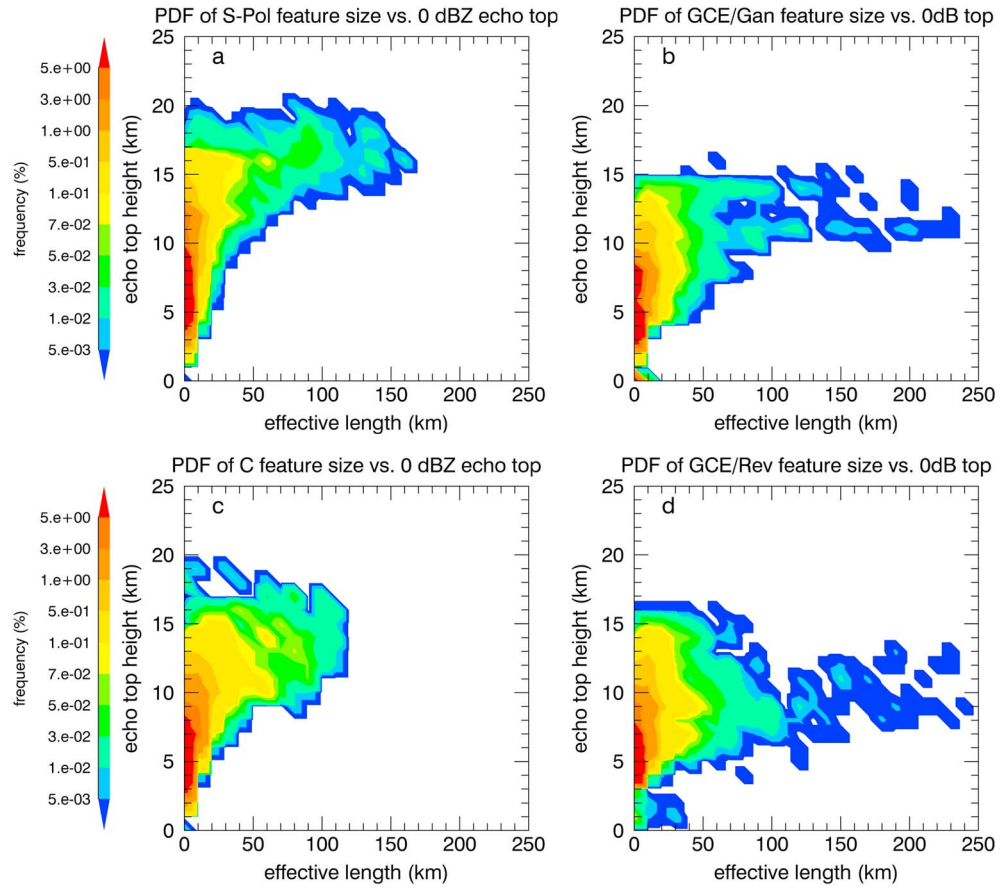


Figure 7. Contour plots of probability density function of the 17-dBZ feature length versus the maximum 0-dBZ echo-top height for every individual precipitation feature. (a) The S-PolKa radar observation. (b) The Goddard Cumulus Ensemble (GCE) model simulation using the Addu large-scale forcing. (c) The C-band radar observation; (d) the GCE simulation using Revelle forcing. Both S-PolKa and C-band use only data within 100 km range. PR = precipitation radar.

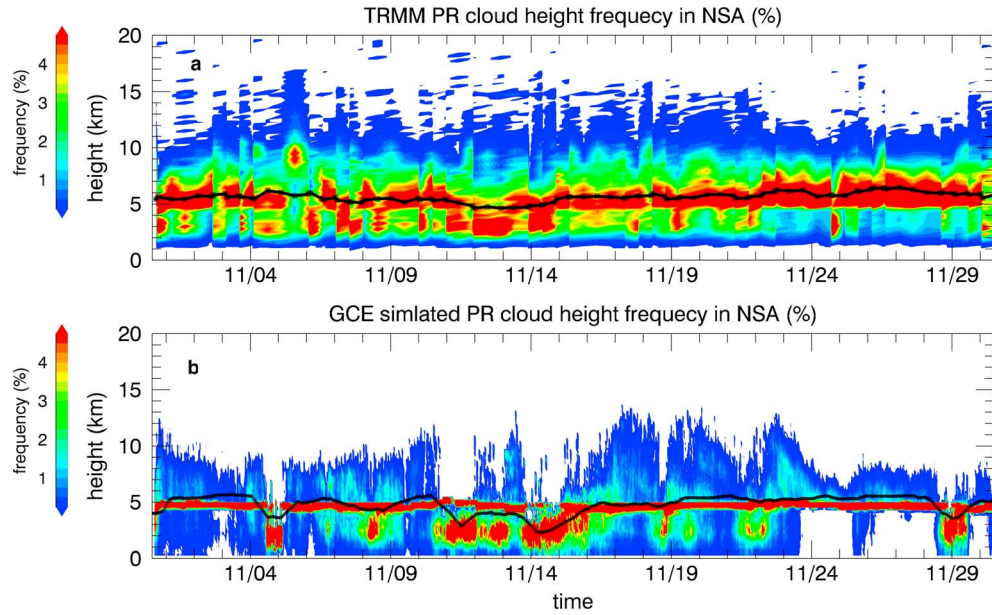


Figure 8. Comparisons of the 17-dBZ radar echo-top height distributions observed by the Tropical Rainfall Measurement Mission precipitation radar (TRMM PR) radar (a) and simulated by the Goddard Cumulus Ensemble (GCE) model using northern sounding array (NSA) forcing (b). The color images are frequencies of the 17-dBZ height occurrences. The black lines represent the median echo-top heights with a 24-hr running mean.

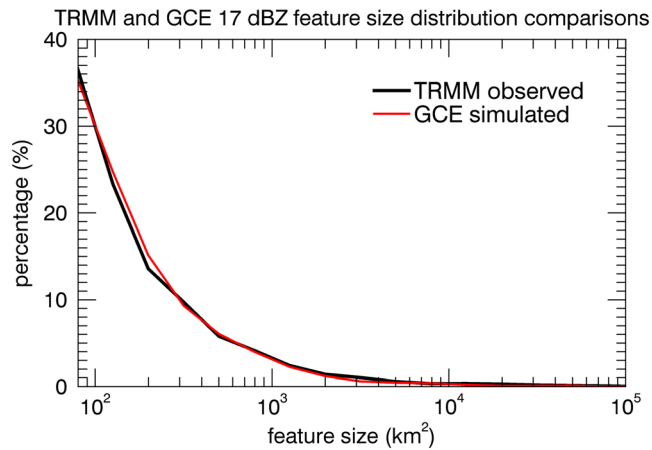


Figure 9. Comparisons of precipitation feature size distributions, defined by the 17-dBZ near surface radar reflectivity. The black line is the Tropical Rainfall Measurement Mission (TRMM) precipitation radar observation between 1 November and 1 December 2011, within the northern sounding array. The red line is the Goddard Cumulus Ensemble (GCE) model simulated feature size distribution over the same period.

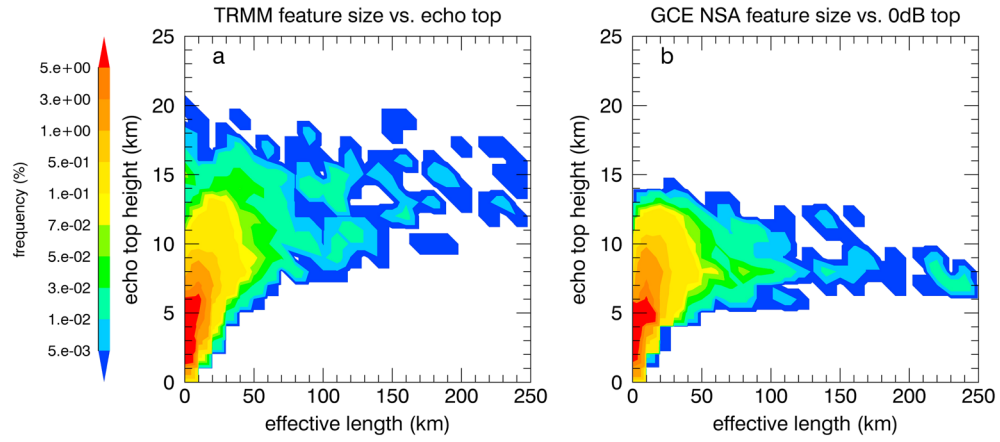


Figure 10. Contours of probability density function of the 17-dBZ precipitation features' effective lengths versus the maximum 17-dBZ echo-top heights. (a) The Tropical Rainfall Measurement Mission (TRMM) precipitation radar observation over the NSA in November 2011. (b) The Goddard Cumulus Ensemble (GCE) simulation using northern sounding array (NSA) forcing over the same period.

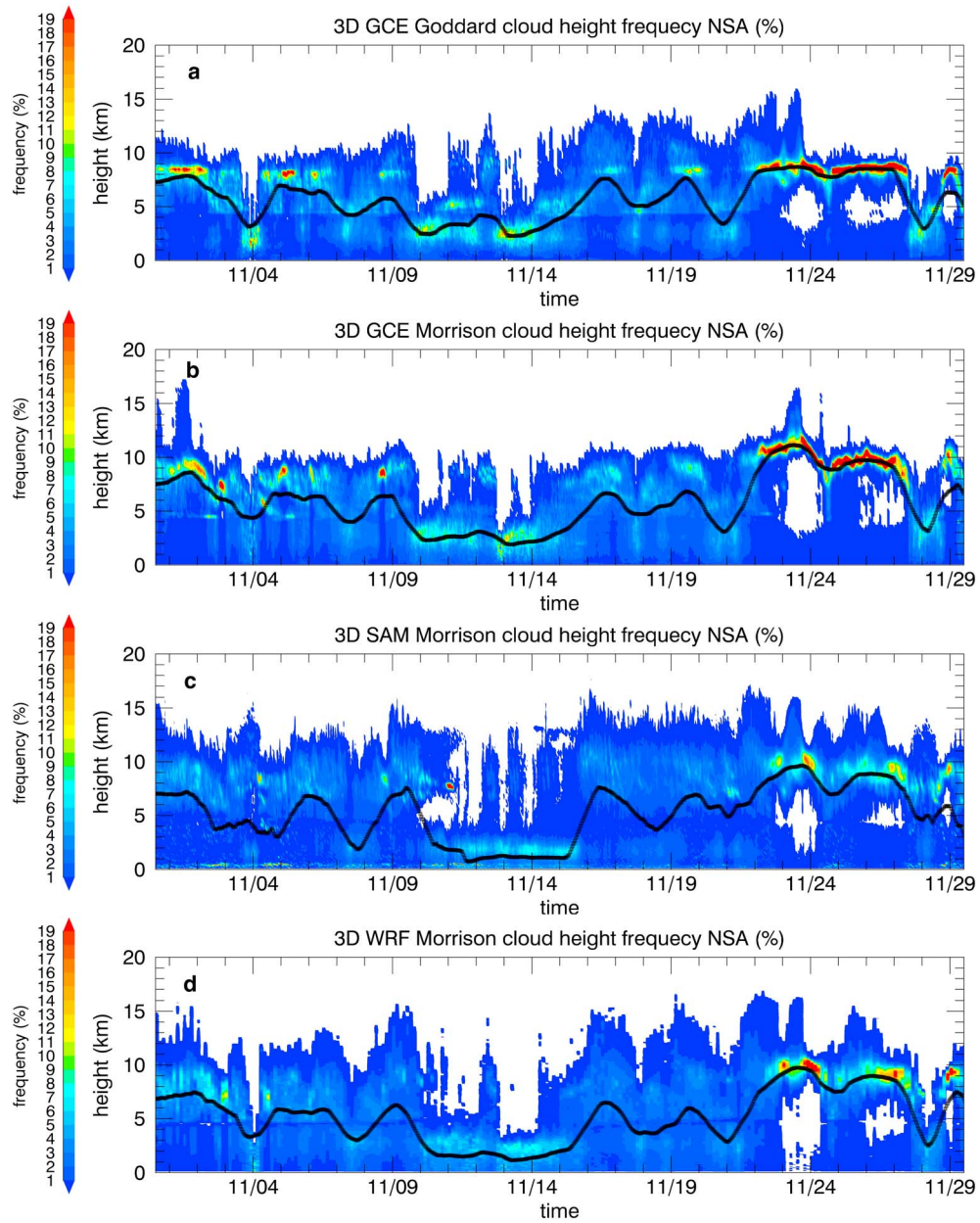


Figure 11. The same as Figure 3 except for cloud-resolving model simulations using the northern sounding array (NSA) forcing for the month of November 2011. (a) The Goddard Cumulus Ensemble (GCE) simulation using the Goddard 3-ICE microphysical scheme; (b) the GCE simulation using the Morrison two-moment scheme; (c) System for Atmospheric Modeling (SAM) and (d) Weather Research and Forecasting (WRF) simulations, both using the Morrison two-moment scheme. The dark lines represent the median cloud top heights with a 24-hr running mean.

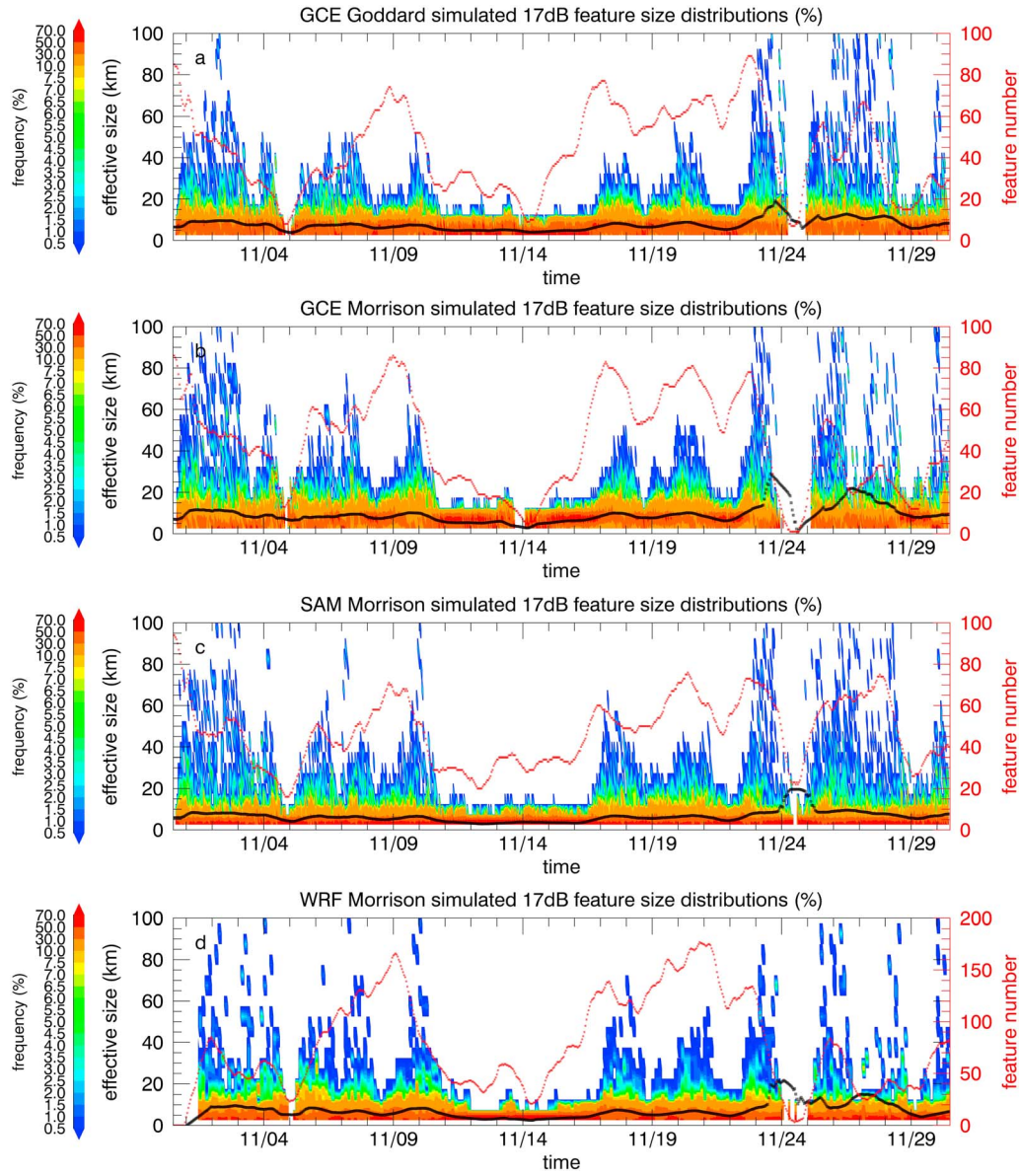


Figure 12. Same as Figure 6 except for different cloud-resolving model simulations. (a) The Goddard Cumulus Ensemble (GCE) model with the 3-ICE microphysical scheme; (b) the GCE model with the Morrison two-moment microphysical scheme; (c) the System for Atmospheric Modeling (SAM) model, also with Morrison scheme; (d) the Weather Research and Forecasting (WRF) model with Morrison scheme.

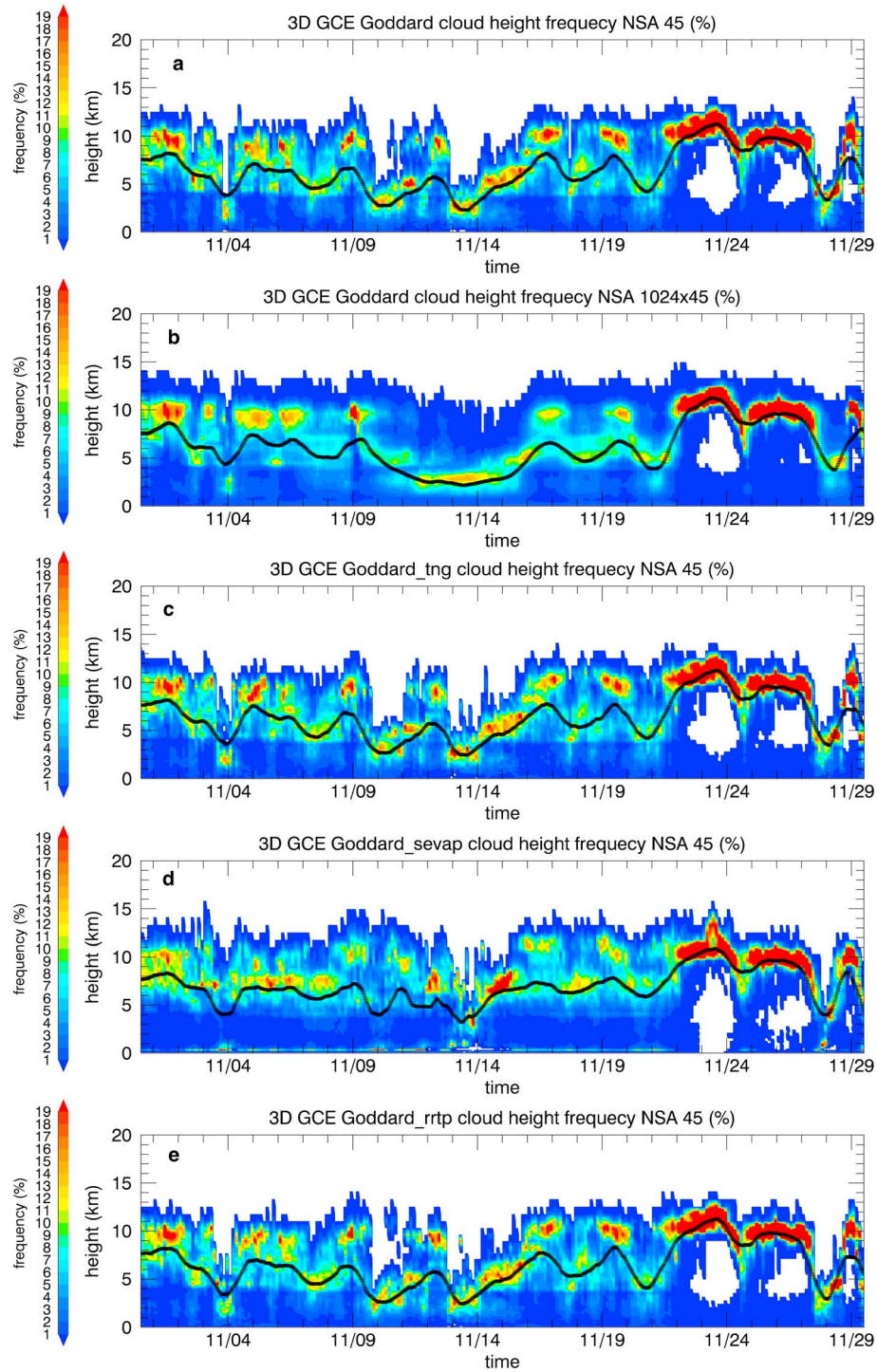


Figure 13. The same as Figure 11 except for Goddard Cumulus Ensemble (GCE) model with Goddard 3-ICE scheme, using coarser vertical resolutions of 45 vertical levels instead of 106 as shown in Figure 11. (a) The small-domain simulation using $256 \text{ km} \times 256 \text{ km}$ horizontal domain with 1 km resolution, as in all previous simulations. (b) Use of $1,024 \text{ km} \times 1,024 \text{ km}$ horizontal domain with 1-km resolution. (c) The same as Figure 13a except for reducing graupel sizes in the Goddard microphysical scheme. (d) The same as Figure 13a except for removing cloud evaporation and ice sublimation where vertical air velocities are higher than 0.5 m/s. (e) The same as Figure 13a except for raising the upper boundary sponge layer 3 km higher. The dark lines again represent the median cloud top heights with a 24-hr running mean. NSA = northern sounding array.

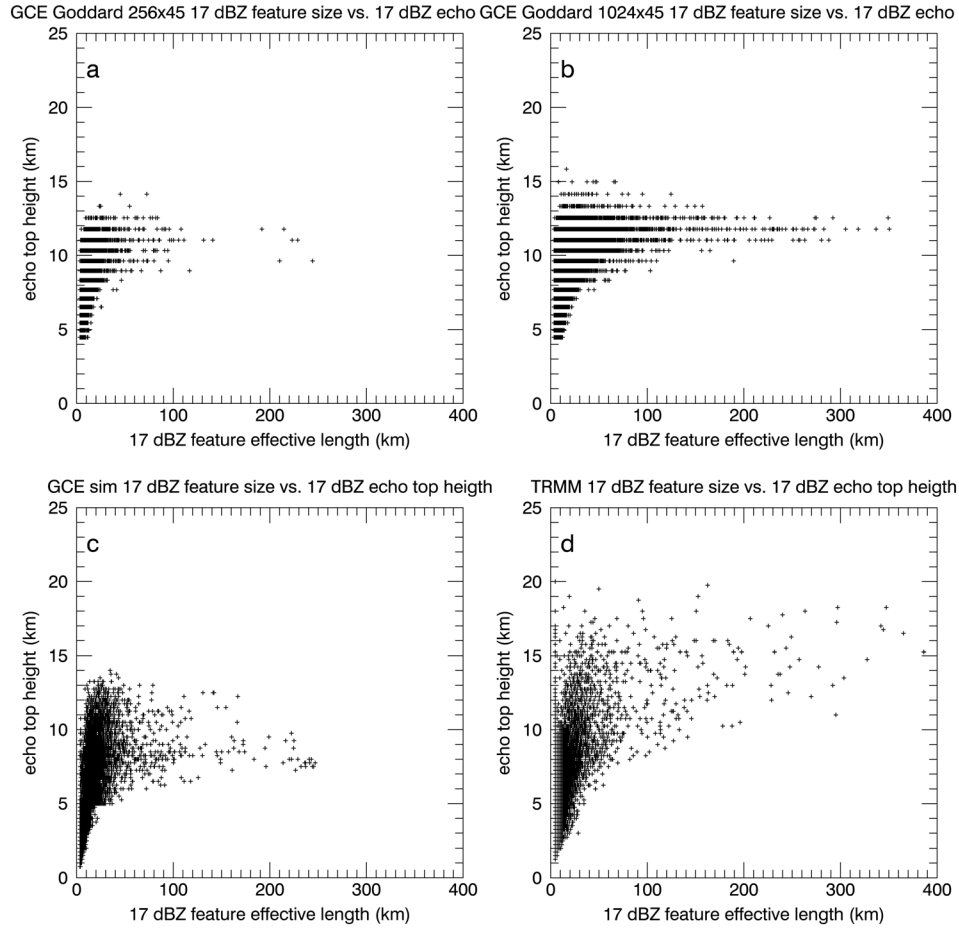


Figure 14. Scatter plots of the precipitation feature effective sizes versus the maximum 17-dBZ echo-top heights, showing feature sizes up to 400 km for large-domain simulation comparisons. (a and c) The small-domain simulations use 256×256 km domain size. Figure 14a has 45 vertical levels, and Figure 14c has 106 vertical levels. (b) The large-domain simulation with $1,024 \times 1,024$ km domain size and 45 vertical levels. (d) The Tropical Rainfall Measurement Mission (TRMM) observation. Figure 14d (Figure 14c) is the same as Figure 10a (Figure 10b) except for an extended x axis range. GCE = Goddard Cumulus Ensemble.

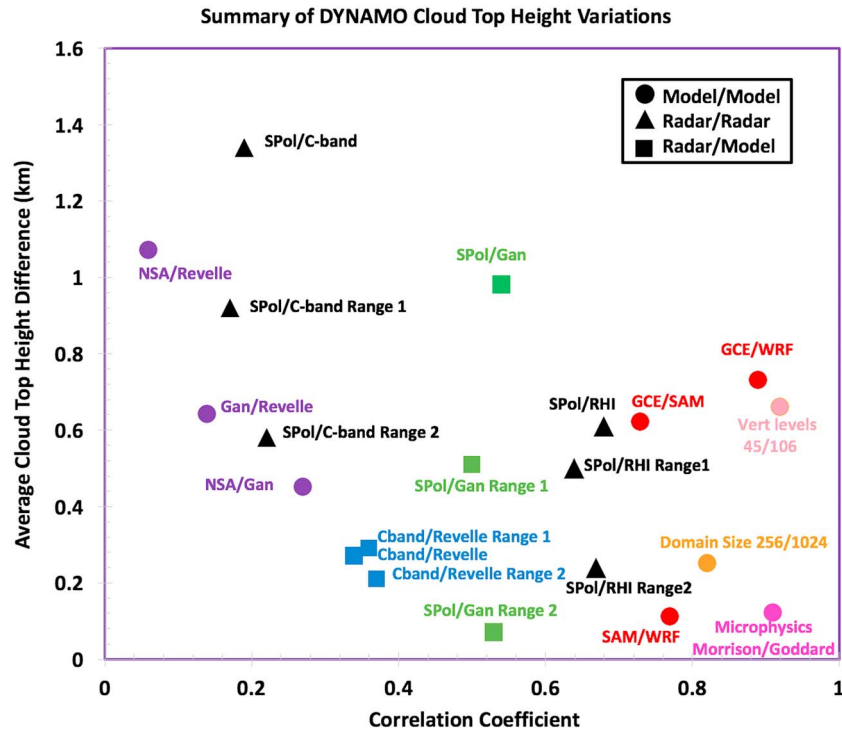


Figure 15. Summary of the radar echo-top height variations during the DYNAMics of the Madden-Julian Oscillation (DYNAMO) November Madden-Julian Oscillation event, using various observations and simulations. The x axis is the correlation coefficients of the median echo-top height time series (the black lines in previous figures); the y axis is their mean differences. Circles represent model-model comparisons; triangles are radar-radar comparisons; squares are model-radar comparisons. Each label indicates the single parameter that differs in the comparison pair, where the first label always has a higher echo-top height. In observations, range 1 represents subsampling of the original data range (0 to 150 km), between 20 and 120 km; range 2 is from 20 to 100 km. GCE = Goddard Cumulus Ensemble; NSA = northern sounding array; WRF = Weather Research and Forecasting; SAM = System for Atmospheric Modeling.